

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:15:58 ; Search time 770.426 Seconds

(without alignments)
1744.014 Million cell updates/sec

Title: US-09-831-272-11

Sequence: 1 tacaatcaacattgttcaacaggaacc 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_srs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_ov:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_srs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_yft:*
38: em_hcg_hum:*
39: em_hcg_mus:*
40: em_hcg_mus:*
41: em_hcg_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	100.0	31	6	AX025611	AX025611 Sequence
2	31	100.0	2683	8	PCPR26	X55736 P.crispum P
3	22.8	73.5	257109	3	AC115577	AC115577 Dictyoste
4	22.6	72.9	224456	10	AC121833	AC121833 Mus muscu
5	22.2	71.6	101110	8	AP006141	AP006141 Lotus cor
6	22	71.0	112695	8	AC119418	AC119418 Medicago
7	22	71.0	130043	8	AC126019	AC126019 Medicago
8	22	71.0	168254	9	AC097502	AC097502 Homo sapi
9	22	71.0	176787	2	AC021374	AC021374 Homo sapi
10	22	71.0	217055	2	AC113221	AC113221 Rattus no
11	22	71.0	221944	2	AC107795	AC107795 Mus muscu
12	22	71.0	300680	2	AC129026	AC129026 Rattus no
13	21.6	69.7	1165	8	YSCMPAG8	M55016 S.italicus
14	21.6	69.7	37245	3	U41036	U41036 Caenorhabd
15	21.6	69.7	41335	3	AC024750	AC024750 Caenorhab
16	21.6	69.7	103608	9	AL139161	AL139161 Human DNA
17	21.6	69.7	165146	2	AC011644	AC011644 Homo sapi
18	21.6	69.7	167163	2	AC013705	AC013705 Homo sapi
19	21.6	69.7	169357	2	AL356462	AL356462 Homo sapi
20	21.6	69.7	185182	9	AC093423	AC093423 Homo sapi
21	21.6	69.7	190599	2	AC146417	AC146417 Pan trogl
22	21.6	69.7	193131	9	AL139076	AL139076 Human DNA
23	21.6	69.7	209365	2	AC006719	AC006719 Caenorhab
24	21.6	69.7	248865	2	AC106672	AC106672 Rattus no
25	21.4	69.0	2000	6	AX356264	AX356264 Sequence
26	21.4	69.0	2000	6	AX819092	AX819092 Sequence
27	21.4	69.0	2000	6	AX830122	AX830122 Sequence
28	21.4	69.0	22131	8	SC9916	Z48952 S.cerevisia
29	21.4	69.0	56510	2	AC020215	AC020215 Drosophila
30	21.4	69.0	66170	2	AC010566	AC010566 Drosophila
31	21.4	69.0	122361	9	HSJ193M13	AL078600 Human DNA
32	21.4	69.0	136131	9	BS000063	BS000063 Pan trogl
33	21.4	69.0	150172	9	AC006285	AC006285 Homo sapi
34	21.4	69.0	166380	9	AC087237	AC087237 Homo sapi
35	21.4	69.0	166490	9	AL359438	AL359438 Human DNA
36	21.4	69.0	167682	2	AC119686	AC119686 Rattus no
37	21.4	69.0	174811	2	AC013259	AC013259 Homo sapi
38	21.4	69.0	180269	9	AC016025	AC016025 Homo sapi
39	21.4	69.0	182477	2	BX649274	BX649274 Danio rer
40	21.4	69.0	184657	3	AC010043	AC010043 Drosophila
41	21.4	69.0	194950	2	AC055115	AC055115 Homo sapi
42	21.4	69.0	208452	2	AC138581	AC138581 Pan trogl
43	21.4	69.0	216302	2	AC107116	AC107116 Rattus no
44	21.4	69.0	241384	2	AC094903	AC094903 Rattus no
45	21.4	69.0	274757	2	BX511000	BX511000 Danio rer

ALIGNMENTS

RESULT 1
AX025611
LOCUS AX025611 31 bp DNA linear PAT 16-SBP-2000
DEFINITION Sequence 11 from Patent WO0029592.
ACCESSION AX025611
VERSION AX025611.1 GI:10187279
KEYWORDS
SOURCE
ORGANISM
Petroselinum crispum (parsley)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
asterids; Campanulids; Apiales; Apiaceae; Apioidae; apioid
superclade; Apium clade; Petroselinum.
REFERENCE
1 Logemann, E., Somseich, I., Hahlbrock, K., Kirsch, C. and Rushton, P.

Pred. No. is the number of results predicted by chance to have a

TITLE Chimeric promoters capable of mediating gene expression in plants
JOURNAL upon pathogen infection and uses thereof
Patent: WO 0029592-A 11 25-MAY-2000;
MAX PLANCK GESSELLSCHAFT (DE) ; LOGSMANN ELKE (DE) ; SOMSICH IMRE
(DE) ; HAHNBROCK KLAUS (DE) ; KIRSCH CHRISTOPH (DE) ; RUSHTON PAUL
(GB)

FEATURES
source location/Qualifiers
1..31
/organism="Petroselinum crispum"
/mol_type="unassigned DNA"
/db_xref="taxon:4043"

ORIGIN
Query Match 100.0%; Score 31; DB 6; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TACAATTCAACATGTTTCAACAGGACC 31
1 TACAATTCAACATGTTTCAACAGGACC 31

Db

RESULT 2
PCPR2G 2683 bp DNA linear PLN 07-JUN-1991
LOCUS P.crispum PR2 gene for pathogenesis-related protein 2.
DEFINITION X55736
ACCESSION X55736.1 GI:20462
KEYWORDS pathogenesis-related protein; PR2 gene.
SOURCE Petroselinum crispum (parsley)
ORGANISM Petroselinum crispum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Apiales; Apiaceae; Apioidae; apioid
superclade; Apium clade; Petroselinum.
1 (bases 501 to 1800)
van de Loch, U., Meier, I., Hahlbrock, K. and Somsich, I. E.
A 125 bp promoter fragment is sufficient for strong
elicitor-mediated gene activation in parsley
EMBO J. 9 (9), 2945-2950 (1990)

JOURNAL 50361009
MEDLINE 2390976
PUBMED 2 (bases 1 to 2683)
AUTHORS Somsich, I. E.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-1991) Somsich I., MAX-PLANCK-INSTITUTE F.
ZUECHTUNGSFORSCHUNG, DEPT. OF BIOCHEMISTRY, CARL-VON-LINNE WEG 10,
5000 KOEHN 30, GERMANY
Evolutionarily conserved polypeptide; close association to disease
resistance in plants.
Origin of genomic library: Douglas et al. EMBO J. 6:189-195 (1987).
location/Qualifiers
1..2683
/organism="Petroselinum crispum"
/mol_type="genomic DNA"
/db_xref="taxon:4043"
/clone="gpcpr2(Pa29)"
/cell_type="cultured parsley cells"
/clone_lib="lambda EMBL4 parsley genomic"
1..790
/evidence=experimental
misc_feature 622..739
/note="elicitor-responsive element"
/evidence=experimental
CAAT_signal 700..704
TATA_signal 759..766
gene 791..1785
/gene="PR2"
mRNA join(791..1077,1273..1785)
/gene="PR2"
/evidence=experimental
prim_transcript 791..1754
/gene="PR2"

exon /evidence=experimental
791..1077
/gene="PR2"
/number=1
join(894..1077,1273..1565)
/gene="PR2"
/codon_start=1
/evidence=experimental
/product="pathogenesis-related protein 2"
/protein_id="CA39268.1"
/db_xref="GI:20463"
/db_xref="GDB:P27538"
/db_xref="SWISS-PROT:P27538"
/translation="MGAVTTDVSVASVPAQITVKGPLDMDNIIPKVLPAIKSIEI
ISGDGAGTIKKVTLGEVSQPTVVKQRIEIDBALKYSYIIEDLILGIIESTSK
FLVVPIDGGCIYKNTITTYPIGDVIVPENVVEATVQSGMVKALVAILNPAAY"
1078..1272
/gene="PR2"
/number=1
1273..1785
/gene="PR2"
/number=2
1749..1754
/gene="PR2"

intron

exon

polya_signal

ORIGIN
Query Match 100.0%; Score 31; DB 8; Length 2683;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TACAATTCAACATGTTTCAACAGGACC 31
715 TACAATTCAACATGTTTCAACAGGACC 745

Db

RESULT 3
AC115577 257109 bp DNA linear INV 12-MAR-2003
LOCUS Dictyostelium discoideum chromosome 2 map 4657875-4914984 strain
DEFINITION AX4, complete sequence.
ACCESSION AC115577.2 GI:28829140
VERSION AC115577
KEYWORDS HTG.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
1 (bases 1 to 257109)
Gloeckner, G., Eichinger, L., Szafranski, K., Pachebat, J., Dear, P.,
Lehmann, R., Baumgart, C., Parra, G., April, J. F., Gulgo, R., Kumpf, K.,
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
Noegel, A. A.
Sequence and analysis of chromosome 2 of Dictyostelium discoideum
Nature 418 (6893), 79-85 (2002)

JOURNAL 22092622
MEDLINE 12097910
PUBMED The Dictyostelium Genome Sequencing Consortium
REMARK 2 (bases 1 to 257109)
REFERENCE Baumgart, C.
AUTHORS Direct Submission
JOURNAL Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
3 (bases 1 to 257109)
Baumgart, C.
JOURNAL Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
4 (bases 1 to 257109)
Baumgart, C.
JOURNAL Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
On Mar 4, 2003 this sequence version replaced gi:1956889.
CDS predictions from GeneID do not necessarily reflect true genes.


```
/codon_start=1
/product="hypothetical protein"
/protein_id="AA051701.1"
/db_xref="GI:28829149"
/translation="MNNFPEWDFKAAARGTGKENTDKYSQMLTLTSLNYLIPPLVPI
NSCKKEPPIINNSVCGKYPVFSGKRDKPKVIGMVGIDIDIDLEVALNELDVLI
DHFFLESTVTHYHKKKPLIMHVFQDRFIKFOKRVHLLIDIDENGKGLPDAS
SYOETRMOKFLYMNERTNOYGDIDIGFDGDEISARLNLHKKCOFRKSDIYDI
GIWEPYGINOVKPKPYSVPMNPYTGDPPTPTIKRASKTKYAPSRNRTSGHYMG
MATHYGYIPQWVKTLSTECGSKNETHVKNSDDIQGNIOKLELRLETKSDHAY
KIQDISTWDFEKEVALIPWFYCNKNRPVWERKNDRLP"
/complete(join(19753..20363,20450..21625,21828..21885))
/note="Genid exon scores (in order of location ranges) :
42.45, 78.19, -3.29 - GSCJ_ID dd_03249"
/codon_start=1
/product="similar to Plasmodium falciparum. Hypothetical
protein"
/protein_id="AA051702.1"
```

Query Match 73.5%; Score 22.8; DB 3; Length 257109;
Best Local Similarity 92.3%; Pred. No. 75;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TACAAATTCAAACATGTTCAACAG 26
Db 228995 TACAAATTCAAACATGTTTAAACAAG 229020

RESULT 4
AC121833 224456 bp DNA linear ROD 27-NOV-2003
LOCUS Mus musculus BAC clone RP24-69C16 from chromosome 3, complete
DEFINITION sequence.

AC121833 GI:30425604
VERSION AC121833.3
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1
AUTHORS Abbot, S., Haakenson, W. and Schatzkammer, K.
TITLE The sequence of Mus musculus BAC clone RP24-69C16
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 224456)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 224456)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 224456)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 224456)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (08-MAY-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

REFERENCE 6 (bases 1 to 224456)
AUTHORS Wilson, R.
TITLE Direct Submision
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2003 this sequence version replaced gi:22475741.

REFERENCE 7
AUTHORS Wilson, R.
TITLE Direct Submision
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2003 this sequence version replaced gi:22475741.

REFERENCE 8
AUTHORS Wilson, R.
TITLE Direct Submision
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2003 this sequence version replaced gi:22475741.

REFERENCE 9
AUTHORS Wilson, R.
TITLE Direct Submision
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2003 this sequence version replaced gi:22475741.

REFERENCE 10
AUTHORS Wilson, R.
TITLE Direct Submision
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 8, 2003 this sequence version replaced gi:22475741.

COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>

Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BB0069C16

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCT-24 BAC Library has been constructed by Pieter de Jong and
coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at <http://www.chori.org>
NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC123057.

FEATURES

source	Location/Qualifiers
1..224456	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="3"
	/map="3"
	/clone="RP24-69C16"
	/clone_11b="RPCT-24"
5..193	/rpt_family="L1"
406..462	/rpt_family="ERVK"
468..659	/rpt_family="B2"
668..792	/rpt_family="A1u"
883..1005	/rpt_family="B4"
1568..2199	/rpt_family="B4"
2281..2427	/rpt_family="L1"
2431..2700	/rpt_family="MER1_type"
2797..2891	/rpt_family="B4"
2906..2961	/rpt_family="B4"
2962..3019	/rpt_family="ID"
3081..3418	/rpt_family="B4"
3476..3538	/rpt_family="MaLR"
3526..3824	/rpt_family="RMR19A"
3859..5045	/rpt_family="RMR19B"
5252..5431	/rpt_family="ERVK"


```

repeat_region      /rpt_family="B2"
                    5523..5669
repeat_region      /rpt_family="Alu"
                    5674..5857
repeat_region      /rpt_family="ERVK"
                    6045..6188
repeat_region      /rpt_family="Alu"
                    6325..6531
repeat_region      /rpt_family="B2"
                    6582..6648
repeat_region      /rpt_family="Alu"
                    7934..8267
repeat_region      /rpt_family="ERVK"
                    8361..8539
repeat_region      /rpt_family="B2"
                    8713..8764
repeat_region      /rpt_family="ID"
                    8890..9036
repeat_region      /rpt_family="MER1_type"
                    9041..9228
repeat_region      /rpt_family="B2"
                    9239..9367
repeat_region      /rpt_family="B4"
                    9443..9584
repeat_region      /rpt_family="Alu"
                    9586..9715
repeat_region      /rpt_family="L1"
                    9718..10040
repeat_region      /rpt_family="L1"
                    10042..10125
repeat_region      /rpt_family="RMR6A"
                    10126..10271
repeat_region      /rpt_family="Alu"
                    10272..10410
repeat_region      /rpt_family="RMR6A"
                    10382..10480
repeat_region      /rpt_family="RMR6A"
                    10481..10655
repeat_region      /rpt_family="B2"
                    10656..10758
repeat_region      /rpt_family="RMR6A"
                    10937..11402
repeat_region      /rpt_family="ERVJ"
                    11769..11941
repeat_region      /rpt_family="L1"
                    12103..12296
repeat_region      /rpt_family="L1"
                    12321..12509
repeat_region      /rpt_family="L1"
                    12571..12708
repeat_region      /rpt_family="Alu"
                    12977..13082
repeat_region      /rpt_family="MALR"
                    13083..13234
repeat_region      /rpt_family="Alu"
                    13235..13267
repeat_region      /rpt_family="MALR"
                    13405..13576
repeat_region      /rpt_family="L2"
                    13746..13840
repeat_region      /rpt_family="Alu"
                    14289..15623
misc_feature       /note="CpG island (%GC=73.8, o/e=0.88, #CpGs=177)"
                    16567..16757
repeat_region      /rpt_family="B2"
                    /product="tRNA-Ser"
                    /note="likely pseudogene (HMM Sc=32.94 / Sec struct
                    Sc=-11.27)"
repeat_region      16886..16988
                    /rpt_family="B4"
repeat_region      17335..17444
                    /rpt_family="MIR"

```

```

repeat_region      17661..17811
                    /rpt_family="Alu"
repeat_region      17832..17932
                    /rpt_family="Alu"
repeat_region      18209..18355
                    /rpt_family="Alu"
repeat_region      18377..18558
                    /rpt_family="B2"
repeat_region      18852..18931
                    /rpt_family="MER1_type"
repeat_region      18956..19071

Query Match       72.9%; Score 22.6; DB 10; Length 224456;
Best Local Similarity 86.2%; Pred. No. 93;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      2 ACATTCAAACATTGTTCAACAGGAC 30
        |||||
Db      113314 ACTGTTCAACCTGTTCAAAAAGGAC 113286

RESULT 5
AP006141      101110 bp      DNA      linear      PLN 22-JUL-2003
LOCUS      Lotus corniculatus var. japonicus genomic DNA, chromosome 4,
DEFINITION      clone:UJT01N13, TM0247, complete sequence.
ACCESSION      AP006141
VERSION      AP006141.1 GI:29122778
KEYWORDS      HTG.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.

REFERENCE
1
AUTHORS      Kaneko,T., Asamizu,E., Kato,T., Sato,S., Nakamura,Y. and Tabata,S.
TITLE      Structural analysis of a Lotus japonicus genome. II. Sequence
features and mapping of sixty-two TAC clones which cover the 6.7 Mb
regions of the genome
JOURNAL      DNA Res. 10 (1), 27-33 (2003)
MEDLINE      22579290
PUBMED      12693552
REFERENCE      2 (bases 1 to 101110)
AUTHORS      Sato,S.
TITLE      Direct Submission
JOURNAL      Submitted (12-DEC-2002) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
Fax:81-438-52-3934)

FEATURES
source
1..101110
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34505"
/chromosome="4"
/clone="LJT01N13"
/clone_1ib="LJT library"
/note="TAC clone:TM0247-synonym: Lotus japonicus"

ORIGIN
Query Match       71.6%; Score 22.2; DB 8; Length 101110;
Best Local Similarity 88.9%; Pred. No. 1,6e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      3 CAATTCAAACATTGTTCAACAGGAA 29
        |||||
Db      93667 CAATTCACAAATTGTTCAACAGGAA 93693

RESULT 6

```

LOCUS	AC119418/c	112695 bp	DNA	linear	PLN 29-APR-2003
DEFINITION	Medicago truncatula clone mth1-23116, complete sequence.				
ACCESSION	AC119418				
VERSION	AC119418.5	GI:30172592			
KEYWORDS	HTG.				
SOURCE	Medicago truncatula (barrel medic)				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; Rosidae; eurosida I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.				
REFERENCE	1 (bases 1 to 112695)				
AUTHORS	Shauli,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.				
TITLE	Medicago truncatula BAC Clone mth1-23116				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 112695)				
AUTHORS	Shauli,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-APR-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA				
REFERENCE	3 (bases 1 to 112695)				
AUTHORS	Shauli,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-AUG-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA				
REFERENCE	4 (bases 1 to 112695)				
AUTHORS	Shauli,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-SEP-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA				
REFERENCE	5 (bases 1 to 112695)				
AUTHORS	Shauli,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-OCT-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA				
REFERENCE	6 (bases 1 to 112695)				
AUTHORS	Shauli,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-APR-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA				
COMMENT	On Apr 29, 2003 this sequence version replaced gi:24371410.				
	----- Genome Center				
	Center: Department Of Chemistry And Biochemistry				
	The University Of Oklahoma				
	Center code:UOKNOR				
FEATURES	-----				
source	Location/Qualifiers				
	1..112695				
	/organism="Medicago truncatula"				
	/mol_type="Genomic DNA"				
	/db_xref="taxon:3880"				
	/clone="mth1-23116"				
ORIGIN	/clone_11b="Medicago truncatula BAC library H1"				
Query Match	71.0%; Score 22; DB 8; Length 112695;				
Best Local Similarity	83.3%; Pred. No. 1.8e+02;				
Matches	25; Conservative 0; Mismatches 5; Indels 0; Gaps 0				
1	TACAAATTCGAAACATTGTCGAAACGAGAAC 30				

RESULT 7	
ACI26019/c	DNA linear PLN 22-APR-2003
LOCUS	ACI26019
DEFINITION	Medicago truncatula clone mch2-22p22, complete sequence.
ACCESSION	ACI26019
VERSION	ACI26019.15 GI:30039760
KEYWORDS	HTG;
SOURCE	Medicago truncatula (barrel medic)
ORGANISM	Medicago truncatula Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabcaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE	1 (bases 1 to 130043) Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A. Medicago truncatula BAC Clone mch2-22p22 Unpublished
JOURNAL	2 (bases 1 to 130043) Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A. Direct Submission Submitted (02-JUL-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE	3 (bases 1 to 130043) Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A. Direct Submission Submitted (16-APR-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
JOURNAL	4 (bases 1 to 130043) Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A. Direct Submission Submitted (22-APR-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
COMMENT	On Apr 22, 2003 this sequence version replaced gi:29893702. ----- Genome Center Center: Department Of Chemistry And Biochemistry The University Of Oklahoma Center code:UONXOR ----- Location/Qualifiers 1.130043 /organism="Medicago truncatula" /mol_type="genomic DNA" /db_xref="taxon:3880" /clone="mch2-22p22" /clone_1fb="Medicago truncatula BAC library H2"
FEATURES	
source	
ORIGIN	
Query Match	71.0%; Score 22; DB 8; Length 130043;
Best Local Similarity	83.3%; Pred. 0.1.de+02;
Matches	25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY	1 TACAATTCGAACATTGTCTCAACAGCAGAAC 30
Db	51160 TTCAATTCGAACATTGTCTCAACAGCAGAAC 51131
RESULT 8	
LOCUS	AC097502 168254 bp DNA linear PRI 20-APR-2002
DEFINITION	Homo sapiens BAC clone RP11-340B18 from 4, complete sequence.
ACCESSION	AC097502 AC021465
VERSION	AC097502.3 GI:20143836
KEYWORDS	HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 168254)
AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Suleston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074

REFERENCE 2 (bases 1 to 168254)
AUTHORS Boatright, E., Haglund, K. and Elliott, G.
TITLE The sequence of Homo sapiens BAC clone RP11-340B18
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 168254)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 168254)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 168254)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 12, 2002 this sequence version replaced gi:11647070.

COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0340B18
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Fenggen, B., Tatenoe, M., Catane, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-511B7, the clone sequenced

to the right is RP11-371E22, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-340B18, actual end is at base position 11859 of RP11-371E22.

FEATURES The sequence of AC021465 has been incorporated into AC097502.

source Location/Qualifiers
1. 168254
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone_id="RP11-340B18"
/clone_1lb="RP11-11"
70..551
/rpt_family="L2"
repeat_region 599..841
/rpt_family="MIR"
repeat_region 858..925
/rpt_family="L2"
repeat_region 1282..1588
/rpt_family="ERV1"
repeat_region 1646..1993
/rpt_family="L2"
repeat_region 1989..2543
/rpt_family="L2"
repeat_region 2561..2986
/rpt_family="MALR"
repeat_region 3017..3349
/rpt_family="L2"
repeat_region 3350..3618
/rpt_family="ERV1"
repeat_region 3686..3817
/rpt_family="ERV1"
repeat_region 4099..4446
/rpt_family="MALR"
repeat_region 4533..4703
/rpt_family="L2"
repeat_region 4819..5670
/rpt_family="ERV1"
repeat_region 6274..6451
/rpt_family="ERV1"
repeat_region 6603..6767
/rpt_family="ERV1"
repeat_region 6768..7164
/rpt_family="L2"
repeat_region 7171..7194
/rpt_family="L2"
repeat_region 7195..7487
/rpt_family="Alu"
repeat_region 7493..7630
/rpt_family="Alu"
repeat_region 7632..7798
/rpt_family="L2"
repeat_region 7801..7921
/rpt_family="L2"
repeat_region 10316..10646
/rpt_family="MALR"
repeat_region 11599..12022
/rpt_family="ERV1"
repeat_region 12421..12507
/rpt_family="ERV1"
repeat_region 12559..12627
/rpt_family="L2"
repeat_region 12767..12864
/rpt_family="MIR"
repeat_region 12865..13162
/rpt_family="Alu"
repeat_region 13163..13261
/rpt_family="MIR"
repeat_region 13352..13600
/rpt_family="MIR"
repeat_region 14433..14672

```

repeat_region      /rpt_family="L1"
                    14873..15166
                    /rpt_family="Alu"
repeat_region      15167..15208
                    /rpt_family="L1"
repeat_region      15348..15846
                    /rpt_family="L2"
repeat_region      16186..16368
                    /rpt_family="MERL_cype"
repeat_region      16369..16681
                    /rpt_family="Alu"
repeat_region      16682..16817
                    /rpt_family="MERL_cype"
repeat_region      17980..18009
                    /rpt_family="AT_rich"
repeat_region      19802..19927
                    /rpt_family="L2"
repeat_region      21923..22157
                    /rpt_family="L1"
repeat_region      22158..22444
                    /rpt_family="Alu"
repeat_region      22445..22503
                    /rpt_family="L1"
repeat_region      22928..22958
                    /rpt_family="AT_rich"
repeat_region      23038..23062
                    /rpt_family="AT_rich"
repeat_region      23318..23612
                    /rpt_family="Alu"
repeat_region      23901..24158
                    /rpt_family="L1"
repeat_region      24159..24427
                    /rpt_family="Alu"
repeat_region      24428..26531
                    /rpt_family="L1"
repeat_region      26548..26585
                    /rpt_family="AT_rich"
repeat_region      26710..27498
                    /rpt_family="L1"
repeat_region      27501..27915
                    /rpt_family="MaLR"
repeat_region      28335..28518
                    /rpt_family="MaLR"
repeat_region      28625..29170
                    /rpt_family="L1"
repeat_region      29171..29659
                    /rpt_family="MaLR"
repeat_region      29660..29767
                    /rpt_family="L1"

```

Query Match 71.0%; Score 22; DB 9; Length 168254;
 Best Local Similarity 83.3%; Pred. No. 1.7e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATTCAAACATTGTTCAACAAGAAC 31

Db 9668 ACAGTTCAAACAGTTTCAACAAGAAC 9697

RESULT 9
 AC021374 176787 bp DNA linear HTG 16-MAR-2000
 LOCUS Homo sapiens clone RP11-25N12, WORKING DRAFT SEQUENCE, 15 unordered
 DEFINITION pieces
 AC021374 AC021374.4 GI:7249071
 VERSION AC021374.4
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 176787)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens, clone RP11-25N12
 Unpublished
 2 (bases 1 to 176787)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G., Casale,A.,
 Choepe,I., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearriano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardina,S., Grant,G., Hagos,B., Harford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kam,L., Karas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marguis,N., McEwan,P., McGuirk,A., McKernan,K.,
 McPeeters,R., Meldrum,J., Menes,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
 Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,
 Stojanovic,N., Sudramanian,A., Talamas,D., Testaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2000 this sequence version replaced gi:6939810.
 All repeats were identified using RepeatMasker:
 Smit,A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

Project Information
 Center project name: L4604
 Center clone name: 25.N12

Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960721
 Consensus quality: 165864 bases at least Q40
 Consensus quality: 170217 bases at least Q40
 Consensus quality: 172405 bases at least Q20
 Insert size: 175387; sum-of-contigs
 Insert size: 176000; agarose-fp
 Quality coverage: 4.8 in Q20 bases; agarose-fp
 Quality coverage: 4.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 15 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1	1339: contig of 1339 bp in length
1340	1439: gap of 100 bp
1440	4249: contig of 2810 bp in length
4250	4349: gap of 100 bp
4350	7864: contig of 3515 bp in length
7865	7964: gap of 100 bp
7965	11523: contig of 3559 bp in length
11524	11623: gap of 100 bp
11624	16505: contig of 4882 bp in length
16506	16605: gap of 100 bp
16606	22103: contig of 5458 bp in length
22104	22203: gap of 100 bp
22204	27892: contig of 5689 bp in length
27893	27992: gap of 100 bp
27993	34300: contig of 6108 bp in length
34301	34400: gap of 100 bp
34401	43494: contig of 9094 bp in length
43495	43594: gap of 100 bp

* 43595 55836: contig of 12242 bp in length
 * 55837 55935: gap of 100 bp
 * 55937 72414: contig of 16478 bp in length
 * 72415 72514: gap of 100 bp
 * 72515 95313: contig of 22799 bp in length
 * 95314 95413: gap of 100 bp
 * 95414 117557: contig of 22144 bp in length
 * 117558 117657: gap of 100 bp
 * 117658 143796: contig of 26141 bp in length
 * 143799 143899: gap of 100 bp
 * 143899 176787: contig of 32889 bp in length.

FEATURES

source

1. 176787
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-25N12"
 /clone_1fb="RP11-11 Human Male BAC"
 1. 1339
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left

misc_feature
 /note="assembly_fragment"
 4350..7864
 /note="assembly_fragment"
 7965..11523
 /note="assembly_fragment"
 11624..16505
 /note="assembly_fragment"
 16606..22103
 /note="assembly_fragment"
 22204..27892
 /note="assembly_fragment"
 27993..34300
 /note="assembly_fragment"
 clone_end:T7
 vector_side:left
 34401..43494
 /note="assembly_fragment"
 43595..55836
 /note="assembly_fragment"
 55937..72414
 /note="assembly_fragment"
 72515..95313
 /note="assembly_fragment"
 95414..117557
 /note="assembly_fragment"
 117658..143798
 /note="assembly_fragment"
 143899..176787
 /note="assembly_fragment"

ORIGIN

Query Match 71.0% Score 22; DB 2; Length 176787;
 Best Local Similarity 83.3% Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATTCAAACATGTTCAACAAAGAAC 31
 Db 103158 ACAGTTCAAACAGTTCAAACAGGAAC 103167

RESULT 10
 AC113221 217055 bp DNA linear HTG 19-NOV-2002
 LOCUS Rattus norvegicus clone CH230-240B21, WORKING DRAFT SEQUENCE.
 DEFINITION AC113221
 AC113221.5 GI:25072721
 VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 217055)

Muzny,D,Marie, Metzker,M, Lee, A, Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhammed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Bunay, C., Burch, P., Butrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Crease, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, T., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorusu, L., Louisedge, H., Lozada, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwankweme, O., Okunolu, G., Olarnpungsoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.,
 Puato, M., Quiroz, J., Rachlin, E., Reeves, K., Register, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, P., Uman, K.,
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willison, R., Wlarczyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, R., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Weinstock, G. and Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstock, G. and Gibbs, R. A.

TITLE

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 217055)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23816990.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSGU
Center clone name: CH230-240B21

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 210001 bases at least Q40
Consensus quality: 211814 bases at least Q30
Consensus quality: 213376 bases at least Q20
Estimated insert size: 217709; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 217055: contig of 217055 bp in length.

Location/Qualifiers

1. 217055
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-240B21"
1. 1949
/note="wga_end_extension
clone_end:T7"
3293. 4127
/note="clone_boundary
clone_end:T7
site:
end_sequence: B2106660"

ORIGIN

Query Match 71.0%; Score 22; DB 2; Length 217055;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 ACAATTCAAACTTTGTTCAACAGAAC 31

Db 164275 ACAATTCAAACTTTGTTCAACAGAAC 164304

RESULT 11
AC107795 221944 bp DNA linear HTG 19-DEC-2003
LOCUS AC107795
DEFINITION Mus musculus chromosome 1 clone RP23-433P9 map 1, *** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
AC107795
VERSION AC107795.7 GI:40217727
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 221944)
Birren, B., Nusbaum, C. and Lander, E.

TITLE JOURNAL REFERENCE AUTHORS

Mus musculus chromosome 1, clone RP23-433P9
Unpublished
2 (bases 1 to 221944)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Batta, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campio, A., Chang, J., Chazaro, B.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeCarliano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardina, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamares, R.,
Lander, T., Lechoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marguis, N., Mathews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retra, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strause, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggilo, J., Vasilev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL REFERENCE AUTHORS

Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 221944)
Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N.,
Anderson, M., Arachchi, H.M., Batta, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chapel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeCarliano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardina, S.,
Graham, L., Grand-Pierre, N., Hater, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lander, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Mathews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retra, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,
Vasilev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (19-DEC-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 19, 2003 this sequence version replaced gi:39930775.
All repeats were identified using RepeatMasker:
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20488
Center clone name: 433_P_9

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 21153: contig of 21153 bp in length
* 21154 21253: gap of 100 bp
* 21254 26183: contig of 4930 bp in length
* 26184 26283: gap of 100 bp
* 26284 36809: contig of 10526 bp in length
* 36810 36909: gap of 100 bp
* 36910 42063: contig of 5154 bp in length
* 42064 42163: gap of 100 bp
* 42164 55640: contig of 13477 bp in length
* 55641 55740: gap of 100 bp
* 55741 71403: contig of 15663 bp in length
* 71404 71503: gap of 100 bp
* 71504 74916: contig of 3413 bp in length
* 74917 75016: gap of 100 bp
* 75017 78702: contig of 3686 bp in length
* 78703 78802: gap of 100 bp
* 78803 90026: contig of 11224 bp in length
* 90027 90126: gap of 100 bp
* 90127 112551: contig of 22425 bp in length
* 112552 124702: gap of 100 bp
* 124703 124802: gap of 100 bp
* 124803 128082: contig of 3280 bp in length
* 128083 128182: gap of 100 bp
* 128183 197215: contig of 68933 bp in length
* 197216 221944: contig of 24729 bp in length.

```

FEATURES

```

Source
1. 221944
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/map="1"
/clone="RP23-433p9"
/clone_1b="RPC1-23 Female Mouse BAC"

```

ORIGIN

```

Query Match 71.0%; Score 22; DB 2; Length 221944;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Oy 1 TACATTCAAACATTTGTTCAACAGAGAC 30
Db 426 TACATTCAAACATGTCACACAGAGAC 455

```

RESULT 12

```

AC129026 300680 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-43K10, WORKING DRAFT SEQUENCE, 4
DEFINITION
unoriented pieces.

```

ACCESSION

```

AC129026
AC129026.3 GI:25073698
HTG: HTGS PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.

```

```

KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

REFERENCE

AUTHORS

1 (bases 1 to 300680)

Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alibekov, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blych, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, J., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensen, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindarne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, B., Milsavljevic, A., Miner, G., Mijca, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muijasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G., Olarnpasegon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reich, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanan, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Wolley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, D., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished
2 (bases 1 to 300680)

Worley, K. C.

Direct Submission

Submitted (25-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 300680)

Rat Genome Sequencing Consortium.

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23267514.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: KANB

Center clone name: CH230-43K10

Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 212742 bases at least Q40
Consensus quality: 215738 bases at least Q30
Consensus quality: 217892 bases at least Q20
Estimated insert size: 213510; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.bgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 258753: contig of 258753 bp in length
* 258754 258853: gap of unknown length
* 258854 287440: contig of 28587 bp in length
* 287441 287540: gap of unknown length
* 287541 298028: contig of 10488 bp in length
* 298029 298128: gap of unknown length
* 298129 300680: contig of 2552 bp in length.

Location/Qualifiers
1. 300680
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-43K10"
complement(256930..257814)
/note="clone_boundary
clone_end:T7
site:
end sequence: BH265728"
287541..288665
/note="wgs_end_extension
clone_end:T7"

misc_feature
misc_feature
misc_feature

ORIGIN

Query Match 71.0%; Score 22; DB 2; Length 300680;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATTCAAACATTGTTCAACAAGAAC 31
Db 229288 ACAATTCAAATATTGTTCAACAAGAAC 229259

RESULT 13
YSCMFAGB 1165 bp DNA linear PLN 27-APR-1993
LOCUS S.italicus mating factor alpha gene, complete cds.
ACCESSION M55016 M35518 M35519
VERSION M55016.1 GI:171944
KEYWORDS mating factor alpha.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 1165)
Kltada, K. and Hishinuma, F.
Evidence for preferential multiplication of the internal unit in
tandem repeats of the mating factor alpha genes in Saccharomyces
yeasts
Curr. Genet. 13 (1), 1-5 (1988)

JOURNAL
MEDLINE 88194702
PUBMED 2834101
COMMENT Original
FEATURES
source 1. 1165
location/Qualifiers
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"

CDS
/db_xref="taxon:4932"
293..853
/codon_start=1
/product="mating factor alpha"
/protein_id="AA134778.1"
/db_xref="GI:171945"
translation="MRPSIFAVLFAASSALAAVNTTDETAQIPEAVIGYLD
BDDFVAVLPFENSTNGILFINTIATIAAEBSVSLDKRAEAMHWLQLKQPMY
KREADAEAMHWLQLKQQPMYKREADAEAMHWLQLKQPMYKREADAEAMHWLQLK
QPMYKREADAEAMHWLQLKQQPMY"
560..598
/product="mating factor alpha"
623..661
/product="mating factor alpha"
686..724
/product="mating factor alpha"
749..787
/product="mating factor alpha"
812..850
/product="mating factor alpha"

mat_peptide
mat_peptide
mat_peptide
mat_peptide
mat_peptide

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 1165;
Best Local Similarity 85.7%; Pred. No. 8.8e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TACAATTCAAACATTGTTCAACAAGAA 28
Db 220 TACGATTCAGAAATGTTCAACAAGAA 247

RESULT 14
U41036 37245 bp DNA linear INV 21-NOV-2002
LOCUS Caenorhabditis elegans cosmid T14B8, complete sequence.
DEFINITION U41036
ACCESSION U41036
VERSION U41036.1 GI:1086751
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Bukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (bases 1 to 37245)
Waterson, R.
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)

JOURNAL
MEDLINE 99069613
PUBMED 9851916
REFERENCE 2 (bases 1 to 37245)
Halloworth, K.
The sequence of C. elegans cosmid T14B8
Unpublished (2001)
REFERENCE 3 (bases 1 to 37245)
Waterson, R.
Direct Submision
Submitted (20-NOV-1995) Robert Waterson
4 (bases 1 to 37245)
Waterson, R.
Direct Submision
Submitted (25-JUN-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
5 (bases 1 to 37245)
Waterson, R.
Direct Submision
Submitted (21-SEP-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
6 (bases 1 to 37245)
Waterson, R.
Direct Submision
Submitted (23-MAY-2002) Department of Genetics, Washington

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
7 (bases 1 to 37245)

Submitted by: Direct Submission
Submitted (21-NOV-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA

Submitted by: Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: tw@emacode.wustl.edu and jee@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=T14E8;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is F22A3, 200 bp overlap; the 3' cosmid is T28B4, 200 bp overlap. Actual start of this cosmid is at base position 197 of T14E8; actual end is at 37245 of T14E8.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yui Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://worfdb.dcel.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. CERNAS are predicted using the program tRNAscan-SE (Lowe, T.M. and Bddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES
source

gene
CDS

```
1. 37245
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="X"
/clone="T14E8"
5098..7244
/gene="T14E8.2"
join(5098..5309,6255..6324,6437..6539,6612..6784,
6855..7244)
/gene="T14E8.2"
/standard_name="T14E8.2"
/notes="contains similarity to Interpro domain IP0002172
(low density lipoprotein-receptor, class A domain)"
/product="Hypotheetical protein T14E8.2"
/protein_id="AA82384.1"
/db_xref="GI:1086752"
```

gene

CDS

```
/db_xref=" WormBase:T14E8.2"
/translation="MOSLEDKLSPPVLCESQOFTITRGVYLCLCCRLKXVYVILR
KLIINHFASSRDRISNPFVPSSESRNRNRVILLPVYKILLAEKNSLTPR
IHRHRKRAQVPRPLSTSPMLWKNDKCMIPGDMACPKDQDELKVLCEP
SALCDHRDGHAEDEDPHRCMEFKLEDAEMRVAEILVLYONNRKXLOPVVHLPD
KDTGIVELPRGVFVEGLGLPLAKSMYENEDDDVKNKSVSESEENDESEOE
DNDKAKEERERRRQEAARQHSRRRTLHLLQLKASYL"
16577..22356
/gene="T14E8.1"
join(16577..16750,17388..17469,17526..17634,17682..17885,
17931..18181,18229..18317,18367..18582,18627..18733,
18913..19015,19063..19182,19228..19416,19465..19586,
19633..20025,20269..20450,20503..20755,20803..20941,
21038..21204,21760..21912,22149..22356)
/gene="T14E8.1"
/standard_name="T14E8.1a"
/notes="contains similarity to Pfam domain PF00069
(Bukaryotic protein kinase domain); coded for by the
following C. elegans cDNAs: YK346e2.5, YK440e9.5,
YK192e3.3, YK272e1.3, YK346e2.3, YK206e9.3, YK304e4.3,
YK375e.3, YK440e9.3, YK1b10.3, YK21h12.3, YK321e10.3,
YK7b2.5, YK1b10.5, YK14c10.3, YK21h12.5, YK30d4.5,
YK37e3.5, YK192e3.5, YK206e9.5, YK208e1.5, YK259e9.5,
YK272e1.5, YK321e10.5"
/codon_start=1
/product="Hypotheetical protein T14E8.1a"
/protein_id="AA82385.2"
/db_xref="GI:1501814"
/translation="WVIGSSQSAKELTQSSIFRFLVLLCFTSATGQINGKLNG
NCFVSRDRLQNKKNKIVGTGFDLIVVAHNSFOVENNERRTGHSVLDHMPRF
LEKLFPSKEIRFYCDSSSGCLCTYSVSSSCSTPANGBEPKIOETLSSAYKIEFG
QIMLAISEKNEEDPDNPRTWILKRYAODGIVIPRTHADSPFIHNNHLDGPERN
YVFVMTASQIFEEVFLHDQSNKKVTVTVIFACADQADLASKISLVGDQERN
ISSRGTAAYDANDLINIMVFNHNSMNMGRPFKANIETKRFITWSTCOEFSFGS
TAKTRCKYPOIFDQMKVKGCLITRSDESPSLCYVAGGADLMDCOLHTASNS
YKRGWLEDYVNOGELMKRIIPYFGIAISLITDGSYPAAVSGERMDVLRFSSE
SADIRPWRNTNISVCKESIITKKNOLLTYVEGLOSDISCKGLPYMCOTLRQGM
EDPLESCWACADNNAORTTTSSEVSSCKNLKHECPSSNMRIKNNSGFTLVVDGR
ALNPKLACGNCVTVVDSSSIQDTPNDIGDSCKVFLSGMIGDNKNSPPEDY
QOADRGTQDVNKSQVDKGGSPGKAIKAIISWTILIVAIIVYWRNPPRLKT
HRRPFGQRIENEDYGMAGROAOLAGINDVYKFRMRPDLKVDKFLVVDKLP
IDGGHGVYKMYSPSKSLERKCYKLEKISFVEARATMSFEPHNLIKIGV
AIDGSHLPIITITTEYAKGDLSEKLENTKEMIDPEPARADIKGMVNSKRTIH
RLACRNCILDRHLYKIKLDFGCRKVIDETLYIOMHERDIPIKFRPRTISQEGFI
TSDINSGVAVIELEFRGSTPYSNASWLLIPMKESTNLRKPPVCEKLYTDM
LACWKNAPRERPOPSDLVTIIPNVYKMGVRSQLOAGYEVSSRFLISHDPAFP
IIONEMNPPLANCONDTNDKTLAEPSDSPSTAIPOSTPOLISECSETSY"
join(16577..16750,17388..17469,17526..17634,17682..17885,
17931..18181,18229..18317,18367..18582,18627..18733,
18913..19015,19063..19182,19228..19416,19465..19586,
19633..20025,20269..20450,20503..20755,20803..20941,
21038..21204,21760..21901,22149..22199)
/gene="T14E8.1"
/standard_name="T14E8.1b"
/notes="contains similarity to Pfam domain PF00069
(Bukaryotic protein kinase domain); coded for by the
following C. elegans cDNAs: YK896b03.3, YK896b03.5"
/codon_start=1
/product="Hypotheetical protein T14E8.1b"
/protein_id="AA106055.1"
/db_xref="GI:15718663"
/translation="WVIGSSQSAKELTQSSIFRFLVLLCFTSATGQINGKLNG
NCFVSRDRLQNKKNKIVGTGFDLIVVAHNSFOVENNERRTGHSVLDHMPRF
LEKLFPSKEIRFYCDSSSGCLCTYSVSSSCSTPANGBEPKIOETLSSAYKIEFG
QIMLAISEKNEEDPDNPRTWILKRYAODGIVIPRTHADSPFIHNNHLDGPERN
YVFVMTASQIFEEVFLHDQSNKKVTVTVIFACADQADLASKISLVGDQERN
ISSRGTAAYDANDLINIMVFNHNSMNMGRPFKANIETKRFITWSTCOEFSFGS
TAKTRCKYPOIFDQMKVKGCLITRSDESPSLCYVAGGADLMDCOLHTASNS
YKRGWLEDYVNOGELMKRIIPYFGIAISLITDGSYPAAVSGERMDVLRFSSE
SADIRPWRNTNISVCKESIITKKNOLLTYVEGLOSDISCKGLPYMCOTLRQGM
EDPLESCWACADNNAORTTTSSEVSSCKNLKHECPSSNMRIKNNSGFTLVVDGR
```

ALNKPRLNACGTCVTVVSSSIQCDTNPDEVIGDSCKVFLSGNIGDKNYSPEPDY
 OQADRGTDVNDKSGVDKSGSPGKMLAIISVMTILLIYAIYVVRNRPRIKT
 HVRRPPIGGRIBENYDMGMAHQAOLANGDNVYKVRMRSPRLKXDNLVKLDP
 IGGHGVYKAMYSKSPSKSLREKVKCKYKEGKISFVEEARPMSPFDNLIKILGV
 AADDSHLLIITTEYKAKGDKLSFIENVENTIMRDLFFRPAFDIAKGNVMSKREIH
 RDLACNCLLDEHLRVKIDFPGICRVADILETELYVMHERDLEVRFPPEISEQGGI
 TSDIMSGVAVITLFTRGSTPYSNMAMWILLIEMLKSESTNLRKRPYCEKLYTDM
 LACWKANPAPERPOFSDLVITLIPVNVKYMEGYDRSQACGYERVELVTSRPSISNLS
 "

complement(23380..32779)
 /gene="Y14B8.3"
 complement(Join(23380..23408,23458..23704,23753..23932,
 23988..24467,24513..24639,25078..25320,25371..25444,
 25486..25604,25662..25726,27723..27849,27904..28022,

Query Match 69.7%; Score 21.6; DB 3; Length 37245;
 Best Local Similarity 85.7%; Pred. No. 3.5e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACATTCAAACATTGTTCAACAAGGA 29
 Db 16418 AAAATTCAAACATTTTAAAGAAGA 16391

RESULT 15
 AC024750/c 41335 bp DNA linear INV 15-JUL-2003
 LOCUS Caenorhabditis elegans cosmid Y17G9A, complete sequence.
 DEFINITION AC024750
 AC024750 GI:7140301
 VERSION AC024750.1 GI:7140301
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 BUKAYOJA; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 41335)
 WILSON,R.
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 99068613
 9851916
 2 (bases 1 to 41335)
 FULTON,B., COURTNEY,L. and AMES,M.
 The sequence of C. elegans cosmid Y17G9A
 Unpublished (2001)
 3 (bases 1 to 41335)
 WATERTON,R.H.
 Direct Submission
 Submitted (01-MAR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 41335)
 WATERTON,R.
 Direct Submission
 Submitted (28-MAR-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 5 (bases 1 to 41335)
 WATERTON,R.
 Direct Submission
 Submitted (24-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 41335)
 WATERTON,R.
 Direct Submission
 Submitted (21-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 41335)
 WILSON,R.
 Direct Submission
 Submitted (15-JUL-2003) Department of Genetics, Washington

COMMENT

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RO, England
 email: submissions@watson.wustl.edu and jee@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m13 subclone.

For a graphical representation of this clone sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=Y17G9A;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is H06H21, 7000 bp overlap; the 3' clone is T08B6, 200
 bp overlap. Actual start of this clone is at base position 18001 of
 K08F11.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 GeneFinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yui Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CR_INDEX.html) and The C.
 elegans ORFome cloning project (<http://worldb.dfci.harvard.edu/>),
 similarity to other proteins from Blastx analyses with C. briggsae
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans Genbank submissions,
 and personal communications with C. elegans researchers. cDNAs
 are predicted using the program tRNAscan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

1..41335
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="IV"
 /clone="Y17G9A"
 /complement(19722..21548)
 /gene="Y17G9A.2"
 complement(Join(19722..19941,19985..20194,20494..20638,
 20684..20796,21514..21548))
 /gene="Y17G9A.2"
 /standard_name="Y17G9A.2"
 /note="coded for by the following C. elegans cDNAs:
 YK726C11.3, YK726C11.5"
 /codon_start=1
 /product="Hypothetical protein Y17G9A.2"
 /protein_id="AA0437.2"
 /db_xref="GI:21166341"
 /db_xref="Wormbase:Y17G9A.2"
 /translation="WQNRRLDDSCPSQTLVLTLSFLCAFTAFSGAFLLYOLMD
 STSCRLPLPLKMDVMPGNVOKIIDIYIENDMHTNDRFTIIRTELMRPKTDEFY
 IIVYHPTESNNHAFYGNKQDIYVINGKSVVYRRFRPKSGSDSAVNYKMYREYDA
 LKENGIFQETDYSHPKOLSRNPTDSNFTIGISNDLNPVRGANNMARIWGWGYFEI
 PVPDGTTRQEKSKITVVGXK"

gene

CDS

```
gene
complement (22168, .23950)
/gene="Y17G9A.3"
/complement (join(22168, .22378, 22896, .23105, 23438, .23582,
23625, .23728, 23916, .23950))
/gene="Y17G9A.3"
/standard_name="Y17G9A.3"
/notes="contains similarity to Pfam domain PF05075 (Protein
of unknown function (DUF684))"
/codon_start=1
/product="Hypothetical protein Y17G9A.3"
/protein_id="AAFE0439.1"
/db_xref="GI:7331751"
/db_xref=" WormBase:Y17G9A.3"
/translation="MQKNRPLKQFTSNGSLVILVGLMFLASSRGLTFRSQRAS
VSSKLPKLTQYIPKVKKIVDVIENDHVTNDPAVLIKTELMKFDITDVFYIMV
YDNKGETSNTSYGNAEQYITSYNPKCNVIYRSHYRSTSRKRNINEVAKICA
EELKQPDGLENLIKNIEDARFICGLEKNRAVSRANSFGARCGMWNKCLEVDPD
TKLTKGQFILIYGFK"
/complement (24532, .26683)
/gene="Y17G9A.4"
/complement (join(24532, .24754, 25107, .25239, 25290, .25393,
25700, .25891, 25939, .26148, 26287, .26431, 26479, .26591,
26649, .26683))
/gene="Y17G9A.4"
/standard_name="Y17G9A.4"
/notes="contains similarity to Interpro domain IPR002040
(tachykinin)"
/codon_start=1
/product="Hypothetical protein Y17G9A.4"
/protein_id="AAFE0440.1"
/db_xref="GI:7331752"
/db_xref=" WormBase:Y17G9A.4"
/translation="MSQNRRTDCENSPKQCLLILISFLCGLIAPGLTYPIQKME
TENVEVSGIPLEKMDYIPQNVKIVDVIENDHVTNDPAVLIKTELMKFNSTDVY
IMVYDNKGETSNTSYGNAEQYITSYNPKCNVIYRSHYRSTSRKRNINEVAKI
CAEGLNKQPDNDVPKLSLENLSRFLVGLKQDLSIRANSFGRWGPGMDSI
DEVYPSKKQTKGSHSOHFLILALISQGLIILQWIEKYSRPLPLEMDY
IMKKNVKIYDVIENDHVTNDPAVLIKTELMKQPDLSISLEKASEIENSRFIGL
FEKIQEVAISANSFNGRWGPGMWNKCLEVYDPSKKTKGQFTLIAGFQ"
27032, .27098
/product="tRNA-His"
/standard_name="Y17G9A.t1"
/notes="C. elegans predicted tRNA"
/db_xref=" WormBase:Y17G9A.t1"
27462, .29380
/gene="str-173"
/join(27462, .27659, 27933, .28020, 28067, .28145, 28355, .28489,
28759, .29032, 29078, .29255, 29313, .29380)
/gene="str-173"
/standard_name="Y17G9A.1"
/notes="contains similarity to Pfam domain PF01461 (7TM
chemoreceptor); coded for by the following C. elegans
cDNAs: OSTRF078B3_1"
/codon_start=1
/product="Seven tm receptor protein 173"
/protein_id="AAFE0438.2"
/db_xref="GI:3269204"
/db_xref=" WormBase:Y17G9A.1"
/translation="MTPVTPANIKTVIOFISVIFINALSYLIVTKSPKIMGYR
HLMIVFCGSMVPSLDVIVQPSIETYSKFLWVDVKNRSLSMSAVLENNLCGI
GTVVGIATHFIRLALERKGRLEPNCQYLPIWFSIPLACTIVIVTLENFGMP
VNTREYIRNPKDSFNMKMECVYGAAYFPIIDKNGKITSMISFRLGCTLTLLTIP
ISTLVAKSMNVQKLEHGESEFANLQWLYKALIAQTLLPVVFFTPFGCVFL
PAEIECQFMSAPRITLVFAIYPAIDPLNLFVDYRNPISNPLRKCCKPKVEAAT
INSTL"
complement (30257, .32936)
/gene="str-172"
/complement (join(30257, .30342, 30391, .30568, 30620, .30893,
31018, .31276, 32600, .32687, 32739, .32936))
/gene="str-172"
/standard_name="Y17G9A.5"
/notes="contains similarity to Pfam domain PF01461 (7TM
chemoreceptor)"
```

```
/codon_start=1
/product="Seven tm receptor protein 172"
/protein_id="AAFE0441.1"
/db_xref="GI:7331753"
/db_xref=" WormBase:Y17G9A.5"
/translation="MHATTFESIKFVQYTSVIFSWLINSPLIYLITKSTSKMGNYR
HLMIVFCPSILIFSTLIDIVQPNIHYYKSAFPMIMDYKNRGIPLMWAAYLICIMGCF

Query Match      69.7%; Score 21.6; DB 3; Length 41335;
Best Local Similarity 85.7%; Pred. No. 3.4e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TACAAATTCAAAGCTGTGTCACCAAGCA 28
Db      30704 TATTAATTCGAGCAATGTGTGAACCAAGCA 30677

Search completed: July 20, 2004, 00:54:45
Job time : 775.426 secs
```

This Page Blank (user)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:16:33 ; Search time 35.4098 seconds
(without alignments)
470.167 Million cell updates/sec

Title: US-09-831-272-2

Perfect score: 30
Sequence: 1 cacactaatcttcacccagtaacatcgcgc 30Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5H_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19.4	64.7	61	1	US-08-207-547A-30	Sequence 30, Appl
2	19.4	64.7	61	1	US-08-215-082-30	Sequence 30, Appl
3	19.4	64.7	61	2	US-08-702-652-30	Sequence 30, Appl
4	17.6	58.7	521	4	US-09-976-594-1112	Sequence 1112, Ap
5	17.2	57.3	768	5	PCT-US95-10973A-29	Sequence 29, Appl
6	17.2	57.3	768	5	PCT-US95-10973A-30	Sequence 30, Appl
7	17.2	57.3	771	5	PCT-US95-10973A-59	Sequence 59, Appl
8	17.2	57.3	804	2	US-08-356-161-3	Sequence 3, Appl
9	17.2	57.3	804	2	US-08-356-161-4	Sequence 4, Appl
10	17.2	57.3	804	2	US-08-356-161-5	Sequence 5, Appl
11	17.2	57.3	804	2	US-08-356-161-6	Sequence 6, Appl
12	17.2	57.3	804	2	US-08-356-161-7	Sequence 7, Appl
13	17.2	57.3	804	3	US-08-718-904-19	Sequence 19, Appl
14	17.2	57.3	804	3	US-08-718-904-20	Sequence 20, Appl
15	17.2	57.3	804	3	US-08-718-904-21	Sequence 21, Appl
16	17.2	57.3	804	3	US-08-718-904-22	Sequence 22, Appl
17	17.2	57.3	804	3	US-08-718-904-23	Sequence 23, Appl
18	17.2	57.3	804	4	US-09-449-249-19	Sequence 19, Appl
19	17.2	57.3	804	4	US-09-449-249-20	Sequence 20, Appl
20	17.2	57.3	804	4	US-09-449-249-21	Sequence 21, Appl
21	17.2	57.3	804	4	US-09-449-249-22	Sequence 22, Appl
22	17.2	57.3	804	5	PCT-US93-05702-3	Sequence 3, Appl
23	17.2	57.3	804	5	PCT-US93-05702-4	Sequence 4, Appl
24	17.2	57.3	804	5	PCT-US93-05702-5	Sequence 5, Appl
25	17.2	57.3	804	5	PCT-US93-05702-6	Sequence 6, Appl
26	17.2	57.3	804	5	PCT-US93-05702-7	Sequence 7, Appl
27	17.2	57.3	804	5	PCT-US93-05702-7	Sequence 7, Appl

c 28	17.2	57.3	804	5	PCT-US95-10973A-3	Sequence 3, Appl
c 29	17.2	57.3	804	5	PCT-US95-10973A-4	Sequence 4, Appl
c 30	17.2	57.3	804	5	PCT-US95-10973A-5	Sequence 5, Appl
c 31	17.2	57.3	804	5	PCT-US95-10973A-6	Sequence 6, Appl
c 32	17.2	57.3	804	5	PCT-US95-10973A-7	Sequence 7, Appl
c 33	17.2	57.3	1167	5	PCT-US95-10973A-57	Sequence 57, Appl
c 34	17.2	57.3	1212	5	PCT-US95-10973A-11	Sequence 31, Appl
c 35	17.2	57.3	1230	2	US-08-356-161-12	Sequence 12, Appl
c 36	17.2	57.3	1230	2	US-08-356-161-13	Sequence 13, Appl
c 37	17.2	57.3	1230	3	US-08-718-904-53	Sequence 53, Appl
c 38	17.2	57.3	1230	3	US-09-449-249-53	Sequence 53, Appl
c 39	17.2	57.3	1230	5	PCT-US93-05702-12	Sequence 12, Appl
c 40	17.2	57.3	1230	5	PCT-US93-05702-13	Sequence 13, Appl
c 41	17.2	57.3	1230	5	PCT-US95-10973A-12	Sequence 12, Appl
c 42	17.2	57.3	1230	5	PCT-US95-10973A-13	Sequence 13, Appl
c 43	17.2	57.3	1251	3	US-08-718-904-74	Sequence 74, Appl
c 44	17.2	57.3	1251	4	US-09-449-249-74	Sequence 74, Appl
c 45	17.2	57.3	1260	3	US-08-718-904-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-08-207-547A-30
; Sequence 30, Application US/08207547A
; Patent No. 5624824
; GENERAL INFORMATION:
; APPLICANT: Yuan, Yan
; APPLICANT: Guerrier-Takeda, Cecilia
; APPLICANT: Altman, Sidney
; APPLICANT: Liu, Remyong
; TITLE OF INVENTION: Targeted Cleavage of RNA Using
; Eukaryotic Ribonuclease P and External Guide Sequence
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,547A
; FILING DATE: 07-MAR-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/03961
; FILING DATE: 28-APR-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/054,892
; FILING DATE: 29-APR-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/931,837
; FILING DATE: 18-AUG-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/875,099
; FILING DATE: 28-APR-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/568,834
; FILING DATE: 17-AUG-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/328,368
; FILING DATE: 24-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: YU100CIP(4)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)-873-8794
TELEFAX: (404)-873-8795
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: tRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-207-547A-30

Query Match 64.7%; Score 19.4; DB 1; Length 61;
Best Local Similarity 58.6%; Pred. No. 2.4;
Matches 17; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACACCTAATTGACCGAGTAACATTCGCC 30
DB 16 ACACUAAUUGCAGCAGAGGUGCGCC 44

RESULT 2
US-08-215-082-30
Sequence 30, Application US/08215082
Patent No. 5728521
GENERAL INFORMATION:
APPLICANT: Yuan, Yan
APPLICANT: Guerrier-Takada, Cecilia
APPLICANT: Altman, Sidney
APPLICANT: Liu, Fenyong
TITLE OF INVENTION: Targeted Cleavage of RNA Using
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patricia L. Pabst
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-4530
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215, 082
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: YU100CIP(4)
FILING DATE: 07-MAR-1994
APPLICATION NUMBER: US PCT/US93/03961
FILING DATE: 28-APR-1993
APPLICATION NUMBER: US 08/054, 892
FILING DATE: 29-APR-1993
APPLICATION DATA:
APPLICATION NUMBER: US 07/931, 837
FILING DATE: 18-AUG-1992
APPLICATION DATA:
APPLICATION NUMBER: US 07/875, 099
FILING DATE: 28-APR-1992
APPLICATION DATA:
APPLICATION NUMBER: US 07/568, 834
FILING DATE: 17-AUG-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/328, 368
FILING DATE: 24-MAR-1989

ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patricia L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: YU100CIP(5)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)-815-6508
TELEFAX: (404)-815-6555
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: tRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-215-082-30

Query Match 64.7%; Score 19.4; DB 1; Length 61;
Best Local Similarity 58.6%; Pred. No. 2.4;
Matches 17; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACACCTAATTGACCGAGTAACATTCGCC 30
DB 16 ACACUAAUUGCAGCAGAGGUGCGCC 44

RESULT 3
US-08-702-652-30
Sequence 30, Application US/08702652
Patent No. 5869248
GENERAL INFORMATION:
APPLICANT: Yan Yuan, Cecilia Guerrier-Takada, and
APPLICANT: Sidney Altman
TITLE OF INVENTION: TARGETED CLEAVAGE OF RNA USING
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patricia L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702, 652
FILING DATE: No. 5869248ember 6, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02816
FILING DATE: March 7, 1995
CLASSIFICATION: 435
APPLICATION DATA:
APPLICATION NUMBER: 08/207, 547
FILING DATE: March 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patricia L.
REGISTRATION NUMBER: 31,284
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)-873-8794
TELEFAX: (404)-873-8795
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```
OTHER INFORMATION: /product= "SAP CYS +10"
FEATURE:
```

NAME/KEY: CDS
 LOCATION: 7..768
 OTHER INFORMATION: /product= "mature SAP CYS +10"
 PCT-US95-10973A-30

Query Match 57.3%; Score 17.2; DB 5; Length 768;
 Best Local Similarity 86.4%; Pred. No. 44;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TAATTGACCGAGTAACATTCG 28
 |||||
 Db 272 TAATATGCCGATTAACATTCG 251

RESULT 7
 PCT-US95-10973A-59/C
 Sequence 59, Application PC/TUS9510973A

GENERAL INFORMATION:
 APPLICANT: Pritz Pharmaeaceuticals, Inc.
 TITLE OF INVENTION: CONJUGATES OF VASCULAR ENDOTHELIAL GROWTH FACTOR WITH TARGET
 NUMBER OF SEQUENCES: 107
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10973A
 FILING DATE: 29-AUG-1995

CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Nottenburg, Carol

REGISTRATION NUMBER: 39,317
 REFERENCE/DOCKET NUMBER: 760100.413PC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 59:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 771 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: both

MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 4..771

OTHER INFORMATION: /product= "SAP CYS -1"
 PCT-US95-10973A-59

Query Match 57.3%; Score 17.2; DB 5; Length 771;
 Best Local Similarity 86.4%; Pred. No. 44;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TAATTGACCGAGTAACATTCG 28
 |||||
 Db 275 TAATATGCCGATTAACATTCG 254

RESULT 8
 US-08-356-161-3/C
 Sequence 3, Application US/08356161
 Patent No. 5916772

GENERAL INFORMATION:
 APPLICANT: Lappi, Douglas A.
 APPLICANT: Barthelemy, Isabel

APPLICANT: Baird, J. Andrew
 APPLICANT: Sosnowski, Barbara A.
 TITLE OF INVENTION: RECOMBINANT PRODUCTION OF
 TITLE OF INVENTION: SAPORIN-CONTAINING PROTEINS
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 701 Fifth Avenue, 6300 Columbia Center
 CITY: Seattle
 STATE: WA
 COUNTRY: USA
 ZIP: 98104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/356,161
 FILING DATE:
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/356,161
 FILING DATE: 13-APR-1995

ATTORNEY/AGENT INFORMATION:
 NAME: No. 5916772cendburg, Carol
 REGISTRATION NUMBER: 39,317

REFERENCE/DOCKET NUMBER: 760100.404US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 804 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown

MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..804

FEATURE:
 NAME/KEY: misc feature
 LOCATION: 1..804

OTHER INFORMATION: /note= "Nucleotide sequence
 OTHER INFORMATION: corresponding to the clone M13 mp18-G4 in Example 1.B.2."

FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: 46..804
 OTHER INFORMATION: /product= "Saporin"

US-08-356-161-3
 Query Match 57.3%; Score 17.2; DB 2; Length 804;
 Best Local Similarity 86.4%; Pred. No. 44;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TAATTGACCGAGTAACATTCG 28
 |||||
 Db 311 TAATATGCCGATTAACATTCG 290

RESULT 9
 US-08-356-161-4/C
 Sequence 4, Application US/08356161
 Patent No. 5916772

GENERAL INFORMATION:
 APPLICANT: Lappi, Douglas A.
 APPLICANT: Barthelemy, Isabel
 APPLICANT: Baird, J. Andrew
 APPLICANT: Sosnowski, Barbara A.
 TITLE OF INVENTION: RECOMBINANT PRODUCTION OF
 TITLE OF INVENTION: SAPORIN-CONTAINING PROTEINS
 NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 701 Fifth Avenue, 6300 Columbia Center
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,161
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/356,161
FILING DATE: 13-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: No. 5916772cemburg, Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.404US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 804 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..804
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..804
OTHER INFORMATION: /note= "Nucleotide sequence
OTHER INFORMATION: corresponding to the clone M13 mp18-cl in Example 1.B.2."
FEATURE:
NAME/KEY: mat.peptide
LOCATION: 46..804
OTHER INFORMATION: /product= "Saporin"
US-08-356-161-4
Query Match 57.3%; Score 17.2; DB 2; Length 804;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 7 TAATTGACCGAGTAACATTCG 28
DB 311 TAATATGCCGATTAACATTCG 290
RESULT 10
US-08-356-161-5/c
Sequence 5, Application US/08356161
Patent No. 5916772
GENERAL INFORMATION:
APPLICANT: Lappi, Douglas A.
APPLICANT: Barthelemy, Isabel
APPLICANT: Baird, J. Andrew
APPLICANT: Sosnowski, Barbara A.
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF
TITLE OF INVENTION: SAPORIN-CONTAINING PROTEINS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 701 Fifth Avenue, 6300 Columbia Center
CITY: Seattle
STATE: WA

COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,161
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/356,161
FILING DATE: 13-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: No. 5916772cemburg, Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.404US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 804 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..804
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..804
OTHER INFORMATION: /note= "Nucleotide sequence
OTHER INFORMATION: corresponding to the clone M13 mp18-G2 in Example 1.B.2."
FEATURE:
NAME/KEY: mat.peptide
LOCATION: 46..804
OTHER INFORMATION: /product= "Saporin"
US-08-356-161-5
Query Match 57.3%; Score 17.2; DB 2; Length 804;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 7 TAATTGACCGAGTAACATTCG 28
DB 311 TAATATGCCGATTAACATTCG 290
RESULT 11
US-08-356-161-6/c
Sequence 6, Application US/08356161
Patent No. 5916772
GENERAL INFORMATION:
APPLICANT: Lappi, Douglas A.
APPLICANT: Barthelemy, Isabel
APPLICANT: Baird, J. Andrew
APPLICANT: Sosnowski, Barbara A.
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF
TITLE OF INVENTION: SAPORIN-CONTAINING PROTEINS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 701 Fifth Avenue, 6300 Columbia Center
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,161
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/356,161
FILING DATE: 13-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: No. 5916772lemburg, Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.40*US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 804 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..804
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..804
OTHER INFORMATION: /note="Nucleotide sequence
corresponding to the clone M13 mp18-G7 in Example 1.B.2."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 46..804
OTHER INFORMATION: /product= "Saporin"
US-08-356-161-6

```

```

OY      7  TAATTGACCGAGTACATTCG  28
      || || || || || || || || || ||
Db      311  TAATATGCCGATTACATTCG  290

RESULT 12
US-08-356-161-7/c
; Sequence 7, Application US/08356161
; Patent No. 5916772
GENERAL INFORMATION:
APPLICANT: Lapol, Douglas A.
APPLICANT: Barthelemy, Isabel
APPLICANT: Baird, J. Andrew
APPLICANT: Sosnowski, Barbara A.
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF
TITLE OF INVENTION: SAVORIN-CONTAINING PROTEINS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 701 Fifth Avenue, 6300 Columbia Center
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,161
FILING DATE:

```

```

1 CLASSIFICATION: 435
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: US 08/356,161
4 FILING DATE: 13-APR-1995
5 ATTORNEY/AGENT INFORMATION:
6 NAME: No. 5916772tenburg, Carol
7 REGISTRATION NUMBER: 39,317
8 REFERENCE/DOCKET NUMBER: 760100.404US
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (206) 622-4900
11 TELEFAX: (206) 682-6031
12 INFORMATION FOR SEQ ID NO: 7:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 804 base pairs
15 TYPE: nucleic acid
16 STRANDEDNESS: double
17 TOPOLOGY: unknown
18 MOLECULE TYPE: cDNA
19 FEATURE:
20 NAME/KEY: CDS
21 LOCATION: 1..804
22 FEATURE:
23 NAME/KEY: misc feature
24 LOCATION: 1..804
25 OTHER INFORMATION: /note= "Nucleotide sequence
26 OTHER INFORMATION: corresponding to the clone M13 mp18-g9 in Example 1.B.2."
27 FEATURE:
28 NAME/KEY: mat peptide
29 LOCATION: 46..804
30 OTHER INFORMATION: /product= "Saporin"
31 US-08-356-161-7

```

```

oy      7  TAATTGACCGAGTAACATTG  28
      ||| ||| ||| ||| ||| |||
Db      311  TAATATGCCGATTACATTG  290

RESULT 13
US-08-718-904-19/c
; Sequence 19, Application US/08718904
; Patent No. 6037329
; GENERAL INFORMATION:
; APPLICANT: Baird, J. Andrew
; APPLICANT: Chandler, Lois Ann
; APPLICANT: Sosnowski, Barbara A.
; TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,904
; FILING DATE: 24-SEP-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6037329tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 760100.415C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

```

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 804 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..804
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..804
OTHER INFORMATION: /note= "Nucleotide sequence
OTHER INFORMATION: corresponding to the clone M13 mp18-G4 in Example I.B.2."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 46..804
OTHER INFORMATION: /product= "Saporin"
US-08-718-904-19

Query Match 57.3%; Score 17.2; DB 3; Length 804;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TAATTGACCGAGTAACTTCG 28
DB 311 TAATATGCCGATTAACTTCG 290

RESULT 14
US-08-718-904-20/c
Sequence 20, Application US/08718904
Patent No. 6037329
GENERAL INFORMATION:
APPLICANT: Baird, J. Andrew
APPLICANT: Sosnowski, Barbara A.
TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
NUMBER OF SEQUENCES: 128
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,904
FILING DATE: 24-SEP-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6037329tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.415C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 804 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS

LOCATION: 1..804
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..804
OTHER INFORMATION: /note= "Nucleotide sequence
OTHER INFORMATION: corresponding to the clone M13 mp18-G1 in Example I.B.2."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 46..804
OTHER INFORMATION: /product= "Saporin"
US-08-718-904-20

Query Match 57.3%; Score 17.2; DB 3; Length 804;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TAATTGACCGAGTAACTTCG 28
DB 311 TAATATGCCGATTAACTTCG 290

RESULT 15
US-08-718-904-21/c
Sequence 21, Application US/08718904
Patent No. 6037329
GENERAL INFORMATION:
APPLICANT: Baird, J. Andrew
APPLICANT: Sosnowski, Barbara A.
TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
NUMBER OF SEQUENCES: 128
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,904
FILING DATE: 24-SEP-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6037329tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.415C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 804 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..804
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..804
OTHER INFORMATION: /note= "Nucleotide sequence
OTHER INFORMATION: corresponding to the clone M13 mp18-G2 in Example I.B.2."
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 46..804
OTHER INFORMATION: /product= "Saporin"

US-08-718-904-21

Query Match 57.3%; Score 17.2; DB 3; Length 804;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TAATTGACCGAGTAACATTG 28
|||||
db 311 TAATATGCCCGATTAAACATTG 290

Search completed: July 20, 2004, 00:55:57
Job time : 36.4098 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:23:29 ; Search time 166.721 Seconds
(Without alignments)
877.864 Million cell updates/sec

Title: US-09-831-272-2

Perfect score: 30
Sequence: I cacacttaattgaccgagtaacatccgc 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCRUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	64.7	3673778	15	US-10-312-841-2
2	19	63.3	684707	16	US-10-398-221-9
3	19	63.3	3011308	16	US-10-398-221-2058
4	18.6	62.0	180	13	US-10-085-783A-42095
5	18.6	62.0	180	16	US-10-242-535A-42095
6	18.6	62.0	612	13	US-10-027-632-217354
7	18.6	62.0	612	13	US-10-027-632-217355
8	18.6	62.0	612	16	US-10-027-632-217354
9	18.6	62.0	612	16	US-10-027-632-217355
10	18.6	62.0	868	16	US-10-350-9238-5
11	18.6	62.0	1033	15	US-10-198-846-9689
12	18.4	61.3	7500	15	US-10-311-455-1088
13	18.4	61.3	33112	16	US-10-429-873A-3
14	18.4	61.3	39001	16	US-10-199-199-14

15	18.4	61.3	65608	9	US-09-962-436-292	Sequence 292, App
16	18.4	61.3	65608	9	US-09-962-832-119	Sequence 119, App
17	18.4	61.3	65608	9	US-09-954-531-180	Sequence 180, App
18	18.2	60.7	49507	13	US-10-087-192-1087	Sequence 1087, App
19	18	60.0	5544	13	US-10-221-613-404	Sequence 404, App
20	18	60.0	5544	15	US-10-240-488-174	Sequence 174, App
21	18	60.0	237961	17	US-10-433-580-2	Sequence 62131, A
22	17.8	59.3	517	13	US-10-027-632-62131	Sequence 62131, A
23	17.8	59.3	517	16	US-10-027-632-62131	Sequence 297953, A
24	17.8	59.3	524	13	US-10-027-632-297953	Sequence 297953, A
25	17.8	59.3	524	16	US-10-027-632-297953	Sequence 310, App
26	17.8	59.3	540	13	US-10-424-599-310	Sequence 226492, A
27	17.8	59.3	654	13	US-10-027-632-226492	Sequence 226492, A
28	17.8	59.3	654	13	US-10-027-632-226492	Sequence 226493, A
29	17.8	59.3	654	16	US-10-027-632-226492	Sequence 26965, A
30	17.8	59.3	654	16	US-10-027-632-226493	Sequence 3712, App
31	17.8	59.3	657	17	US-10-437-963-26965	Sequence 7, Appli
32	17.8	59.3	893	15	US-10-198-846-3712	Sequence 3814, App
33	17.8	59.3	1049	15	US-10-198-846-11907	Sequence 1, Appli
34	17.8	59.3	10597	15	US-10-057-108-10	Sequence 53539, A
35	17.8	59.3	10597	17	US-10-738-886-10	Sequence 53539, A
36	17.8	59.3	10599	15	US-10-057-108-11	Sequence 614, App
37	17.8	59.3	10599	17	US-10-738-886-11	Sequence 142, App
38	17.8	59.3	12482	15	US-10-057-108-7	
39	17.8	59.3	12482	17	US-10-738-886-7	
40	17.8	59.3	198285	9	US-09-880-107-3814	
41	17.8	59.3	3673778	15	US-10-312-841-2	
42	17.6	58.7	291	13	US-10-027-632-53539	
43	17.6	58.7	291	16	US-10-027-632-53539	
44	17.6	58.7	440	16	US-10-291-265-614	
45	17.6	58.7	478	16	US-10-291-265-142	

ALIGNMENTS

RESULT 1
US-10-312-841-2/C
Sequence 2, Application US/10312841
Publication No. US20030186277A1
GENERAL INFORMATION:
APPLICANT: Epigenomics AG
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des HNC
FILE REFERENCE: E01/1208/KO
CURRENT APPLICATION NUMBER: US/10/312, 841
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 2
SEQ ID NO 2
LENGTH: 3673778
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (379615)
US-10-312-841-2

Query Match 64.7%; Score 19.4; DB 15; Length 3673778;
Best Local Similarity 79.3%; Pred. No. 6.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACACCTTAATTGACGAGTAACCTTTGCC 30
DB 2679336 ACACCAAAATTAACCAATTAACATGACC 2679308

RESULT 2
US-10-398-221-9/C
Sequence 9, Application US/10398221
Publication No. US20040018514A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederick

```
APPLICANT: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 684707
TYPE: DNA
ORGANISM: Listeria innocua
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(end)
OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9
```

```
Query Match      63.3%; Score 19; DB 16; Length 684707;
Best Local Similarity 81.5%; Pred. No. 8.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      2  ACGCTTAATTTGACCGAGTAACATTCG 28
DB      7378 ACGCTTAATTTGACCGATTAACATTCG 7352
```

```
RESULT 3
US-10-398-221-2058/c
Sequence 2058, Application US/10398221
Publication No. US20040018514A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederik
APPLICANT: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2058
LENGTH: 3011208
TYPE: DNA
ORGANISM: Listeria innocua
US-10-398-221-2058
```

```
Query Match      63.3%; Score 19; DB 16; Length 3011208;
Best Local Similarity 81.5%; Pred. No. 8.5e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      2  ACGCTTAATTTGACCGAGTAACATTCG 28
DB      2444961 ACGCTTAATTTGACCGATTAACATTCG 2444935
```

```
RESULT 4
US-10-085-783A-42095
Sequence 42095, Application US/10085783A
Publication No. US20040037841A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
CURRENT FILING DATE: 2002-02-28
```

```
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 42095
LENGTH: 180
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4)..(4)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (6)..(6)
OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-42095
```

```
Query Match      62.0%; Score 18.6; DB 13; Length 180;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      2  ACGCTTAATTTGACCGAGTAACATT 26
DB      20  AACCTTAATTTGACCGAGTACCTTT 44
```

```
RESULT 5
US-10-242-535A-42095
Sequence 42095, Application US/10242535A
Publication No. US2004001363A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 42095
LENGTH: 180
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4)..(4)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (6)..(6)
OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-42095
```

```
Query Match      62.0%; Score 18.6; DB 16; Length 180;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      2  ACGCTTAATTTGACCGAGTAACATT 26
DB      20  AACCTTAATTTGACCGAGTACCTTT 44
```

```
RESULT 6
US-10-027-632-217354/c
; Sequence 217354, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 217354
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-217354

Query Match          62.0%; Score 18.6; DB 13; Length 612;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 CACTTAATTGACCGAGTAACATTC 27
Db      381 CATTTAATTGACGAGTAACACATC 357

RESULT 7
US-10-027-632-217355/c
; Sequence 217355, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 217355
; LENGTH: 612
```

```
TYPE: DNA
ORGANISM: Human
US-10-027-632-217355

Query Match          62.0%; Score 18.6; DB 13; Length 612;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 CACTTAATTGACCGAGTAACATTC 27
Db      381 CATTTAATTGACGAGTAACACATC 357

RESULT 9
US-10-027-632-217355/c
; Sequence 217355, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
```

```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 217355
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-217355
```

```

Query Match      62.0%; Score 18.6; DB 16; Length 612;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      3 CACTTAATTGACCGAGTACATTTC 27
DB      381 CATTTAATTGACGATATACACTC 357
```

```

RESULT 10
US-10-350-923B-5
; Sequence 5, Application US/10350923B
; Publication No. US2004002084A1
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Ma, Xiao Jun
; TITLE OF INVENTION: Nucleic acid, polypeptides, vectors, and cells derived from activ
; TITLE OF INVENTION: eosinophil cells
; FILE REFERENCE: S03157-01
; CURRENT APPLICATION NUMBER: US/10/350, 923B
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: US/ 09/454, 280
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: PCT/US99/28773
; PRIOR FILING DATE: 1999-06-12
; PRIOR APPLICATION NUMBER: US 60/111,006
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 868
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (34)..(37)
; OTHER INFORMATION: n is selected from the group consisting of a, c, g, and t
US-10-350-923B-5
```

```

Query Match      62.0%; Score 18.6; DB 16; Length 868;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      2 ACCTTAATTGACCGAGTACATT 26
DB      694 AAACCTTAATTGACGAGTACCTTT 718
```

```

RESULT 11
US-10-198-846-9689
; Sequence 9689, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steilmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
```

```

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9689
; LENGTH: 1033
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1_2_3_4_5_6_7_8_9_10_11_1028_1029_1030_1031,
; LOCATION: 1032, 1033
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9689
```

```

Query Match      62.0%; Score 18.6; DB 15; Length 1033;
Best Local Similarity 84.0%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      2 ACCTTAATTGACCGAGTACATT 26
DB      763 AAACCTTAATTGACGAGTACCTTT 787
```

```

RESULT 12
US-10-311-455-1088/c
; Sequence 1088, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1088
; LENGTH: 7500
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1088
```

```

Query Match      61.3%; Score 18.4; DB 15; Length 7500;
Best Local Similarity 78.6%; Pred. No. 6.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY      3 CACTTAATTGACCGAGTACATTGCC 30
DB      4616 CACTTAATAACGACGAGTACATTGCC 4589
```

```

RESULT 13
US-10-429-873A-3
; Sequence 3, Application US/10429873A
; Publication No. US20040014193A1
; GENERAL INFORMATION:
; APPLICANT: ABUTHREIDER, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
```


;; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;; TITLE OF INVENTION: THEREOF
;; FILE REFERENCE: CL001364
;; CURRENT APPLICATION NUMBER: US/10/429,873A
;; CURRENT FILING DATE: 2003-05-06
;; NUMBER OF SEQ ID NOS: 8
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 33112
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; US-10-429-873A-3

Query Match 61.3%; Score 18.4; DB 16; Length 33112;
Best Local Similarity 78.6%; Pred. No. 8.4e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACACTTAATTGACCGAGTAACATTGCG 29
DB 23043 ACCCTTTTCTGACCGAGTAACATTGAC 23070

RESULT 14
US-10-199-199-14
; Sequence 14, Application US/10199199
; Publication No. US20040014047A1
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowsett
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF LIM DOMAIN KINASE 1 EXPRESSION
; FILE REFERENCE: RTS-0375
; CURRENT APPLICATION NUMBER: US/10/199,199
; CURRENT FILING DATE: 2002-07-18
; NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 14
; LENGTH: 39001
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-199-199-14

Query Match 61.3%; Score 18.4; DB 16; Length 39001;
Best Local Similarity 78.6%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACACTTAATTGACCGAGTAACATTGCG 29
DB 36190 ACCCTTTTCTGACCGAGTAACATTGAC 36217

RESULT 15
US-09-962-436-292
; Sequence 292, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; SOFTWARE: PatentIn version 3.0
; NUMBER OF SEQ ID NOS: 568
; SEQ ID NO 292
; LENGTH: 65608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: n,a,t,g or c

US-09-962-436-292

Query Match 61.3%; Score 18.4; DB 9; Length 65608;
Best Local Similarity 78.6%; Pred. No. 9.7e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACACTTAATTGACCGAGTAACATTGCG 29
DB 36260 ACCCTTTTCTGACCGAGTAACATTGAC 36287

Search completed: July 20, 2004, 01:48:38
Job time : 173.721 secs

and 1000 (up to)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:17:50 ; Search time 1374.59 Seconds

(without alignments)
651.733 Million cell updates/sec

Title: US-09-831-272-2

Perfect score: 30

Sequence: 1 cacactaattgcgcgagtaacatccgc 30

Scoring table: Gapped 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estba:*
16: em_estba:*
17: em_estba:*
18: em_estba:*
19: em_estba:*
20: em_estba:*
21: em_estba:*
22: em_estba:*
23: em_estba:*
24: em_estba:*
25: em_estba:*
26: em_estba:*
27: em_estba:*
28: gb_est1:*
29: gb_est2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	22.2	74.0	660	CF349641	CF349641 fc2569.e
2	20.6	68.7	500	CD581429	CD581429 RK012A4H0
3	20.2	67.3	602	BJ521358	BJ521358 BJ521358
4	20.2	67.3	682	BJ022103	BJ022103 BJ022103

c	5	20.2	67.3	685	12	BJ533932	BJ533932
c	6	20.2	67.3	740	12	BJ015002	BJ015002
c	7	20	66.7	982	10	BF385898	BF385898
c	8	19.6	65.3	253	9	AV084908	AV084908
c	9	19.6	65.3	291	9	AV163173	AV163173
c	10	19.6	65.3	556	9	AI909003	AI909003
c	11	19.6	65.3	606	10	BG080964	BG080964
c	12	19.6	65.3	759	29	EX187708	EX187708
c	13	19.6	65.3	980	28	CC448991	CC448991
c	14	19.4	64.7	389	14	CD981231	CD981231
c	15	19.4	64.7	494	28	AO627538	AO627538
c	16	19.4	64.7	526	9	AA552611	AA552611
c	17	19.4	64.7	619	10	BE539667	BE539667
c	18	19.4	64.7	787	28	CC392698	CC392698
c	19	19.4	64.7	803	28	BZ994426	BZ994426
c	20	19.4	64.7	823	28	CC392700	CC392700
c	21	19.4	64.7	827	28	BZ994430	BZ994430
c	22	19.4	64.7	908	28	BZ722043	BZ722043
c	23	19.2	64.0	306	10	BB208476	BB208476
c	24	19.2	64.0	406	14	CD606005	CD606005
c	25	19.2	64.0	425	28	AQ023481	AQ023481
c	26	19.2	64.0	548	28	AQ579759	AQ579759
c	27	19.2	64.0	574	28	AQ571694	AQ571694
c	28	19.2	64.0	608	10	AM420732	AM420732
c	29	19.2	64.0	1201	13	BX420783	BX420783
c	30	19	63.3	119	12	BM035642	BM035642
c	31	19	63.3	321	14	CD598710	CD598710
c	32	19	63.3	374	14	CD283854	CD283854
c	33	19	63.3	405	14	CB599807	CB599807
c	34	19	63.3	424	12	BM279675	BM279675
c	35	19	63.3	494	10	BF384422	BF384422
c	36	19	63.3	502	12	BG307726	BG307726
c	37	19	63.3	505	12	BG305109	BG305109
c	38	19	63.3	518	13	BQ261533	BQ261533
c	39	19	63.3	520	12	BI672771	BI672771
c	40	19	63.3	540	12	BM530029	BM530029
c	41	19	63.3	556	12	BM530969	BM530969
c	42	19	63.3	614	12	BG307193	BG307193
c	43	19	63.3	614	12	BM775766	BM775766
c	44	19	63.3	649	13	BU710008	BU710008
c	45	19	63.3	654	12	BM660238	BM660238

ALIGNMENTS

RESULT 1	CF349641	660 bp	mRNA	linear	EST 18-AUG-2003
LOCUS	fc2569.e	Rose Petals (Fragrant Cloud) Lambda Zap Express Library			
DEFINITION	Rosa hybrid cultivar cDNA clone fc2569.e 5', mRNA sequence.				
ACCESSION	CF349641				
VERSION	CF349641.1	GI:33832991			
KEYWORDS	EST.				
SOURCE	Rosa hybrid cultivar				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids I; Rosales; Rosaceae; Rosoideae; Rosa.				
REFERENCE	Guterman, I., Shalit, M., Menda, N., Piestun, D., Dafny-Yelin, M., Shalev, G., Bar, E., Davydov, O., Ovadis, M., Emanuel, M., Wang, J., Adam, Z., Pichersky, E., Lewinson, E., Zamir, D., Vainstein, A. and Weiss, D.				
AUTHORS	Rose scent: genomics approach to discovering novel floral fragrance-related genes				
TITLE	Plant Cell 14 (10), 2325-2338 (2002)				
JOURNAL	22256146				
MEDLINE	12368489				
PUBMED	Contact: Naama Menda				
COMMENT	Petal Genomics				
	Faculty of Agricultural, Food and Environmental Quality Sciences,				
	The Hebrew University of Jerusalem				

P.O. Box 12, Rehovot, 76100, Israel
Tel: 972 8 9489 683
Fax: 972 8 9468 265
Email: shaham@agri.huji.ac.il

Seq primer: T3 forward.
Location/Qualifiers

FEATURES

source

1..660
/organism="Rosa hybrid cultivar"
/mol_type="mRNA"
/strain="Pragant Cloud"
/db_xref="taxon:128735"
/clone="fc2569.c"
/issue_type="Petals"
/dev_stage="Young open flower at stage four"
/clone_lib="Rose Petals (Pragant Cloud) Lambda Zap Express Library"
/note="Vector: pBKCW, Site_1: EcoRI, Site_2: XhoI"

ORIGIN

Query Match 74.0%; Score 22.2; DB 14; Length 660;
Best Local Similarity 88.9%; Pred. No. 26;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CACTTAATTGACCGAGTACATTGCG 29
|||||
441 CACTTAATTGACCGAGTACATTGCG 467

RESULT 2
CD581429 500 bp mRNA linear EST 16-JUN-2003
LOCUS RK012A4H08 T3 Zebrafish Kidney Marrow cDNA library Danio rerio cDNA
DEFINITION clone RK012A4H08 5', mRNA sequence.
ACCESSION CD581429
VERSION CD581429.1 GI:31758092
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 500)
Song,H.D., Wu,X.Y., Sun,X.J., Zhou,Y., Liu,T.X., Zhang,G.W., Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y., Look,A.T., Zou,L.I. and Chen,Z.
Gene Expression Profiling in the Zebrafish Kidney Marrow Tissue Unpublished (2003)
Contact: Chen Z.
State Key Lab for Medical Genomics
Shanghai Institute of Hematology, Ruijin Hospital Affiliated to Shanghai Second Medical University
197 Rui Jin Road II, Shanghai 200025, P. R. China
Tel: 86-21-64740490
Fax: 86-21-64743206
Email: zchen@scn.sh.cn
Seq primer: T3.

FEATURES

source

1..500
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="RK012A4H08"
/dev_stage="mature"
/clone_lib="Zebrafish Kidney Marrow cDNA library"
/note="Organ: Kidney; Vector: pBS-CMV; Site_1: XhoI; Site_2: EcoRI; Total RNA was extracted from the kidney tissues of mature zebrafish. The poly (A)+ RNA fraction was separated from total RNA by oligo (dT) cellulose chromatography. Library was initially constructed in the lambdaZAP Express vector (Stratagene) and in vivo excised into pBS-CMV vector."

ORIGIN

Query Match 68.7%; Score 20.6; DB 14; Length 500;
Best Local Similarity 85.2%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CACTTAATTGACCGAGTACATTGCG 29
|||||
22 CACTTCTTTGACCGAGTACATTGCG 48

DB

RESULT 3

LOCUS

BU521358 602 bp mRNA linear EST 09-AUG-2002
BU521358 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB018H19 5',
DEFINITION mRNA sequence.
ACCESSION BU521358
VERSION BU521358.1 GI:22180170
KEYWORDS EST.

SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 602)
Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadasu Shin-i
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES

source

1..602
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF01SSB018H19"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSB cDNA"

ORIGIN

Query Match 67.3%; Score 20.2; DB 12; Length 602;
Best Local Similarity 88.0%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTACAT 25
|||||
DB 228 CTCACCTAATTGACCGAGAGCAT 252

RESULT 4

LOCUS

BU022103 682 bp mRNA linear EST 05-DEC-2001
BU022103 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA118F02 3',
DEFINITION mRNA sequence.
ACCESSION BU022103
VERSION BU022103.1 GI:17371006
KEYWORDS EST.

SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 682)
Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab

JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1. .682
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSA118F02"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_1lb="MF01SSA cDNA"

ORIGIN
Query Match 67.3%; Score 20.2; DB 12; Length 682;
Best Local Similarity 88.0%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACACTTAATTGACCGAGTAACAT 25
Db 440 CTCACCTTAATTGACCGAGTAACAT 416

RESULT 5
BJS33932/c 685 bp mRNA linear EST 09-AUG-2002
LOCUS BJS33932 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB018H19 3',
DEFINITION mRNA sequence.
ACCESSION BJS33932
VERSION BJS33932.1 GI:22192744
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes

REFERENCE
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1. .685
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSB018H19"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_1lb="MF01SSB cDNA"

ORIGIN
Query Match 67.3%; Score 20.2; DB 12; Length 685;
Best Local Similarity 88.0%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACACTTAATTGACCGAGTAACAT 25
Db 1 CACACTTAATTGACCGAGTAACAT 25

Db 461 CTCACCTTAATTGACCGAGTAACAT 437

RESULT 6
BUT015002/c 740 bp mRNA linear EST 07-DEC-2001
LOCUS BUT015002 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA017F10 3',
DEFINITION mRNA sequence.
ACCESSION BUT015002
VERSION BUT015002.1 GI:17406329
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes

REFERENCE
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1. .740
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSA017F10"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_1lb="MF01SSA cDNA"

ORIGIN
Query Match 67.3%; Score 20.2; DB 12; Length 740;
Best Local Similarity 88.0%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CACACTTAATTGACCGAGTAACAT 25
Db 441 CTCACCTTAATTGACCGAGTAACAT 417

RESULT 7
BF385998/c 982 bp mRNA linear EST 27-NOV-2000
LOCUS BF385998 60204617F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4195675 5',
DEFINITION mRNA sequence.
ACCESSION BF385998
VERSION BF385998.1 GI:11367303
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.

RESULT 10
 A1909003/c 556 bp mRNA linear EST 30-MAR-2000

LOCUS A1909003
 DEFINITION QV-BT197-050499-008 BT197 Homo sapiens cDNA, mRNA sequence.
 ACCSSION A1909003
 VERSION A1909003.1 GI:64999683
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663
 PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/seq/gethtml.pl?l=QV&t2=QV-BT197-008.html&t3=050499&t4=1>)
 Seq primer: puc 18 forward.

FEATURES
 source
 Location/Qualifiers
 1..556
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /sex="female"
 /dev_stage="Adult"
 /clone_lib="BT197"
 /note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
 Query Match 65.3%; Score 19.6; DB 9; Length 556;
 Best Local Similarity 84.6%; Pred. No. 3.6e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACGAGTACATT 26
 Db 263 CAGACTTAATTGCCGATTAACATT 238

RESULT 11
 BG080964/c 606 bp mRNA linear EST 17-DEC-2003

LOCUS BG080964
 DEFINITION H3059E08-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3059E08 5', mRNA sequence.
 ACCSSION BG080964
 VERSION BG080964.2 GI:40015143
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Tanaka, T.S., Jardadt, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grabovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, M.H. III, Becker, K.G. and Ko, M.S.H.
 Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray

TITLE

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)

MEDLINE 20381348
 PUBMED 10922068

COMMENT On Jan 26, 2001 this sequence version replaced gi:12563532.
 Other ESTs: H3059E08-3
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please visit <http://igsun.grc.nia.nih.gov/cDNA/15k.html> for details.
 Plate: H3059 row: E column: 08
 Seq primer: -21M13 Reverse
 High quality sequence stop: 606
 POLYA=No.

FEATURES
 source
 Location/Qualifiers
 1..606
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="nlaesr:H3059E08-5"
 /db_xref="taxon:10090"
 /clone="H3059E08"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /clone_lib="NIA Mouse 15K cDNA Clone Set"
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN
 Query Match 65.3%; Score 19.6; DB 10; Length 606;
 Best Local Similarity 84.6%; Pred. No. 3.6e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CACTTAATTGACCGAGTACATTG 28
 Db 37 CACTTAAGTTTACCGAGTACATTG 12

RESULT 12
 BX187708/c 759 bp DNA linear GSS 28-JAN-2003

LOCUS BX187708
 DEFINITION Danio rerio genomic clone DKEX-194P17, genomic survey sequence.
 ACCSSION BX187708
 VERSION BX187708.1 GI:28019594

KEYWORDS GSS.
SOURCE Danilo rerio (zebrafish)
ORGANISM Danilo rerio
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 759)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk Unpublished
This sequence was generated from the T7 end of BAC 194P17, 194P17 is part of the Daniokey BAC library created by R. Plasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_rerio/.

FEATURES
source
1..759
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-194P17"
/issue_type="Testis"
/note="vector pindigobAC-536"

ORIGIN
Query Match 65.3%; Score 19.6; DB 29; Length 759;
Best Local Similarity 84.6%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGATACATT 26
505 CACTTAATATATACCGATACATT 480

RESULT 13
LOCUS CC448991 980 bp DNA linear GSS 22-MAY-2003
DEFINITION ZMMBRC0322H05f ZMMBRC Zea mays subsp. mays genomic clone
ACCESSION CC448991
VERSION CC448991.1 GI:30996554
KEYWORDS GSS.
SOURCE Zea mays subsp. mays (maize)
ORGANISM Zea mays subsp. mays
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 980)
Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C., Rouzard,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.
Sequencing of the maize genome at PGIR (2003b)
Unpublished (2003)
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers University
190 Freilichuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 3.
Location/Qualifiers
1..980
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultiivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBRC0322H05"
/lab_host="E. coli DH10B"

FEATURES
source
1..980
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultiivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBRC0322H05"
/lab_host="E. coli DH10B"

ORIGIN
/clone_lib="ZMMBRC"
/note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"

Query Match 65.3%; Score 19.6; DB 28; Length 980;
Best Local Similarity 84.6%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGATACATT 26
808 CACTTAATTGACCAAGTACATT 833

RESULT 14
LOCUS CD981231/c 389 bp mRNA linear EST 16-JUL-2003
DEFINITION QAI11901.Y9 QAI Zea mays cDNA clone QAI121901, mRNA sequence.
ACCESSION CD981231
VERSION CD981231.1 GI:32841550
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 389)
Genopiante.
Genopiante, a major partnership french program in plant genomics
Unpublished (2003)
Contact: Genopiante
Genopiante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french plant genomics programme "genopiante" (<http://www.genopiante.com> and <http://genopiante-info.infobio.gen.fr>).
Location/Qualifiers
1..389
/organism="Zea mays"
/mol_type="mRNA"
/cultiivar="P2"
/db_xref="taxon:4577"
/clone="QAI121901"
/issue_type="ear leaf"
/clone_lib="QAI"

ORIGIN
Query Match 64.7%; Score 19.4; DB 14; Length 389;
Best Local Similarity 79.3%; Pred. No. 4.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACACTTAATTGACCGATACATTGCC 30
183 ACAAAAGTTTGACCGATACATTCAACC 155

RESULT 15
LOCUS AQ627538 494 bp DNA linear GSS 16-JUN-1999
DEFINITION CITBI-EI-2653P15.TF CITBI-EI Homo sapiens genomic clone 2653P15, genomic survey sequence.
ACCESSION AQ627538
VERSION AQ627538.1 GI:5089930
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 494)
Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.
Use of BAC End Sequences from Caltech Libraries for Sequence-Ready

JOURNAL
COMMENT

Map Building
Unpublished (1997)
Other_GSSs: CIRBI-B1-2653P15.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hde@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.ligr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. .494
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2653P15"
/sex="male"
/cell_type="sperm"
/clone_lib="CIRBI-B1"
/notes="Vector: pBelBAC1; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"

ORIGIN

Query Match 64.7%; Score 19.4; DB 28; Length 494;
Best Local Similarity 79.3%; Pred. No. 4.3e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTACATTCCC 29
DB 303 CACATGTAATGTGTCGAGTACATTCCC 331

Search completed: July 20, 2004, 01:42:45
Job time: 1377.59 secs

175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:17:50 ; Search time 1420.41 Seconds
(without alignments)
651.733 Million cell updates/sec

Title: US-09-831-272-11

Perfect score: 31

Sequence: 1 tacaatcaacatgttcaacaagaagac 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estfun:*
17: em_estfun:*
18: em_estfun:*
19: em_estfun:*
20: em_estfun:*
21: em_estfun:*
22: em_estfun:*
23: em_estfun:*
24: em_estfun:*
25: em_estfun:*
26: em_estfun:*
27: em_estfun:*
28: gb_est1:*
29: gb_est2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.2	74.8	844	28	BZ392622 EINH33TF
2	22	71.0	568	28	AZ109637 RPCI-23-1
3	22	71.0	706	28	BZ676143 PUBB31TD
4	22	71.0	842	28	BZ669683 PUBAV94TD

c	5	21.8	70.3	489	29	CPA561291
	6	21.6	69.7	572	10	BC044802
	7	21.6	69.7	587	12	BQ040901
	8	21.6	69.7	675	10	BE824450
	9	21.6	69.7	687	13	CA130450
	10	21.6	69.7	689	28	AO780940
	11	21.6	69.7	734	28	AZ202685
	12	21.4	69.0	283	9	AV961084
	13	21.4	69.0	407	12	B161768
	14	21.4	69.0	486	13	BM195405
	15	21.4	69.0	486	13	BM209523
	16	21.4	69.0	560	9	AV963578
	17	21.4	69.0	561	13	BM274645
	18	21.4	69.0	581	13	BM277525
	19	21.4	69.0	587	9	AV992847
	20	21.4	69.0	591	13	BM288634
	21	21.4	69.0	612	9	AV999016
	22	21.4	69.0	614	9	AV673090
	23	21.4	69.0	620	13	BM243946
	24	21.4	69.0	623	9	AV951238
	25	21.4	69.0	627	9	AV951117
	26	21.4	69.0	638	13	BM291174
	27	21.4	69.0	665	13	BM213642
	28	21.4	69.0	699	29	BK140539
	29	21.4	69.0	703	12	B1573756
	30	21.4	69.0	737	13	BM272276
	31	21.4	69.0	804	29	CNS0090L
	32	21.4	69.0	891	28	AZ538099
	33	21.4	69.0	944	28	BH151913
	34	21.2	68.4	245	14	CA246963
	35	21.2	68.4	245	14	CA248260
	36	21.2	68.4	304	13	CA149735
	37	21.2	68.4	362	14	CF090483
	38	21.2	68.4	507	14	CA244864
	39	21.2	68.4	527	14	CA189166
	40	21.2	68.4	530	14	CA115205
	41	21.2	68.4	539	14	CA223116
	42	21.2	68.4	584	14	CA240951
	43	21.2	68.4	585	14	CA223520
	44	21.2	68.4	665	28	BH005742
	45	21.2	68.4	673	14	CA233741

ALIGNMENTS

RESULT 1
LOCUS BZ392622 844 bp DNA linear GSS 30-APR-2003
DEFINITION EINH33TF EI_10_12_KB Entamoeba invadens genomic clone EINH33,
genomic survey sequence.
ACCESSION BZ392622
VERSION BZ392622.1 GI:30239159
KEYWORDS GSS.
SOURCE Entamoeba invadens
ORGANISM Entamoeba invadens
REFERENCE Wang,Z., Samuelson,J., Clark,C.G., Eichinger,D., Paul,J., van
Dellen,K., Hall,N., Anderson,I., and Loftus,B.
Gene discovery in the Entamoeba invadens genome
Mol. Biochem. Parasitol. 129 (1), 23-31 (2003)
JOURNAL MEDLINE
PUBMED 12798503
COMMENT Other GSSs: EINH33TR
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-858-3543
Fax: 301-858-0208
Email: enta@tigr.org
DNA was provided by Daniel Eichinger

Seq primer: TF
Class: sheared ends.
Location/Qualifiers

1.844
/organism="Entamoeba invadens"
/mol_type="genomic DNA"
/strain="IP-1"
/db_xref="taxon:33085"
/clone_1lb="E1_10_12_KB"
/note="Vector: pHS2; Site 1: BstXI; Total genomic DNA was isolated from early log phase trophozoites of E. invadens IP-1 using a Qiaagen plant DNA extraction kit. A shotgun medium-size plasmid library (average insert size of 10 - 12 kb) was generated by random mechanical shearing of E. invadens genomic DNA, repairing the ends of DNA fragments with T4 Polymerase, adding BstXI adaptors and ligating into the BstXI site of a pUC-derived vector pHS2."

ORIGIN

Query Match 74.8%; Score 23.2; DB 28; Length 844;
Best Local Similarity 89.3%; Pred. No. 1.7e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 AATTCAACATTGTTCAACAGAAC 31
Db 716 AATTAAACTTAGTCAACAGAAC 743

RESULT 2
AZ109637 568 bp DNA linear GSS 09-MAY-2000
LOCUS RPCI-23-12E1.TV RPCI-23 Mus musculus genomic clone RPCI-23-12E1,
DEFINITION genomic survey sequence.
ACCESSION AZ109637
VERSION AZ109637.1 GI:7762695
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Buekaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 568)
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shattman, S., Akinret, B., Levins, M., McGann, S., Tsagaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)

TITLE
JOURNAL
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

libray available, please contact Peter de Jong
(piederdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (http://www.tigr.org/tcd/bac_ends/mouse/bac_end_intro.html)
plate: 12 row: B column: 1
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers

1.568
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-12E1"
/sex="Female"
/lab_host="DH10B"
/clone_1lb="RPCI-23"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Query Match 71.0%; Score 22; DB 28; Length 568;
Best Local Similarity 83.3%; Pred. No. 4.8e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TACAATTCACATGTTCAACAGAAC 30
Db 417 TCAATGCAAGATGTGACACAGAAC 446

RESULT 3

BZ676143/c 706 bp DNA linear GSS 05-FEB-2003
LOCUS PUBB31TD_ZM_0.6_1.0_KB Zea mays genomic clone ZM8TA014P14,
DEFINITION genomic survey sequence.
ACCESSION BZ676143
VERSION BZ676143.1 GI:28227523
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE

Buekaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 706)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Uteerback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennettzen, J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw

TITLE
JOURNAL
COMMENT Contact: Cathy Whitelaw

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org

Seq primer: TF
Class: sheared ends.

FEATURES
source Location/Qualifiers

1.706
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_1lb="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high Cot selected genomic DNA library"

ORIGIN

Query Match 71.0%; Score 22; DB 28; Length 706;
Best Local Similarity 83.3%; Pred. No. 4.8e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TACAATTCACATGTTCAACAGAAC 30
Db 85 TAAATTTAACTTTTCAACAGAAC 56

RESULT 4

BZ669683 842 bp DNA linear GSS 05-FEB-2003
LOCUS PUBAV94TD_ZM_0.6_1.0_KB Zea mays genomic clone ZM8TA013020,
DEFINITION genomic survey sequence.
ACCESSION BZ669683
VERSION BZ669683.1 GI:28216592

KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 842)
AUTHORS Whitelaw, C.A., Queckenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J., and Bennetzen, J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.

FEATURES
source Location/Qualifiers
1..842
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_1b="ZMBR1A013020"
/clone_1b="ZM_0.6_1.0 KB"
/note="Vector: PCR4-TOPO, Site 1: EcoRI, 0.6-1.0 kb high cor selected genomic DNA library"

ORIGIN
Query Match 71.0%; Score 22; DB 28; Length 842;
Best Local Similarity 83.3%; Pred. No. 4.7e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TACAAATTCACATGTTCTCAACAGGAC 30
Db 46 TAAATTTAAACCTTTTCAACAGGAC 17

RESULT 5
CPA561291/c 489 bp DNA linear GSS 11-AUG-2003
LOCUS CPA561291
DEFINITION Crypsosporidium parvum GSS, PAC clone pica_0001_d10, T7 end
sequence, genomic survey sequence.
ACCESSION AJ561291
VERSION AJ561291.1 GI:31335740
KEYWORDS GSS: genome survey sequence.
SOURCE Crypsosporidium parvum
ORGANISM Crypsosporidium parvum
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Crypsosporidiidae; Crypsosporidium.

REFERENCE 1
AUTHORS Bankier, A.T., Spriggs, H.F., Partmann, B., Konfortov, B.A., Madera, M., Vogel, C., Teichmann, S.A., Ivens, A. and Dear, P.H.
TITLE Integrated mapping, chromosomal sequencing and sequence analysis of Crypsosporidium parvum
JOURNAL Genome Res. 13 (8), 1787-1799 (2003)
MEDLINE 22784085
PUBMED 12869580
REFERENCE 2 (bases 1 to 489)
AUTHORS Dear, P.H.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2003) Dear P.H., PNBC Biotech Division, MRC Laboratory of Molecular Biology, Hills Road, Cambridge, Cambs CB2 2QH, UNITED KINGDOM

FEATURES
source Location/Qualifiers
1..489
/organism="Crypsosporidium parvum"
/mol_type="genomic DNA"
/seqtype="Type 2"
/isolate="Iowa"

/db_xref="taxon:5807"
/clone_1b="pica_0001_d10"
/clone_1b="pica"
/dev_stage="oocyst"
/note="T7 end sequence"

ORIGIN
Query Match 70.3%; Score 21.8; DB 29; Length 489;
Best Local Similarity 92.0%; Pred. No. 5.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ATTCAACATGTTCTCAACAGGAA 29
Db 61 ATTCAACATGTTCTCAACAGTGA 37

RESULT 6
BG044802 572 bp mRNA linear EST 28-NOV-2001
LOCUS BG044802
DEFINITION aa32g02.y1 Gm-c1059 Glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-c1059-700 5', mRNA sequence.
ACCESSION BG044802
VERSION BG044802.1 GI:12491894
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 572)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Matra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Persson, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McGinn, R., Waterston, R. and Wilson, R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Reggen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact via email: cdu@reggen.com
High quality sequence stop: 421.

FEATURES
source Location/Qualifiers
1..572
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone_1b="Gm-c1059"
/tissue_type="whole seedling, 2 week old, etiolated"
/lab_host="DH10B"
/clone_1b="Gm-c1059"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from 2 week old etiolated whole seedlings of P1468916. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed in the laboratory of Dr. Randy Shoemaker at Iowa state university."

ORIGIN

Query Match 69.7%; Score 21.6; DB 10; Length 572;
 Best Local Similarity 85.7%; Pred. No. 6.8e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 AATTCAAACATTGTTCAACAGGAACC 31
 |||||
 Db 256 AATTGAGAAATGTTCAACAGGAACC 283

RESULT 7

LOCUS B0040901 587 bp mRNA linear EST 17-APR-2002
 DEFINITION gdc39h11.y1 Moss EST library PPS Physcomitrella patens cDNA clone
 PEP SOURCE ID:PPS101622 5' similar to TR:P93818 P93818 HYPOTHETICAL
 50.5 KD PROTEIN.;, mRNA sequence.

ACCESSION B0040901 GI:19788753
 VERSION B0040901.1
 KEYWORDS EST.
 SOURCE Physcomitrella patens
 ORGANISM Physcomitrella patens

REFERENCE 1 (bases 1 to 587)
 Quatrano, R., Bashardes, S., Cove, D., Cumig, A., Knight, C.,
 Clifton, S., Marra, M., Hillier, L., Page, D., Martin, J., Wylie, T.,
 Underwood, K., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Smaller, T., Steptoe, M., Gibbons, M., Harvey, N., Ritter, E.,
 Jackson, Y., McCam, R., Waterston, R. and Wilson, R.
 Leada/Mash u Moss EST Project
 Unpublished (1999)
 Other ESTs: gdc39h11.x1
 CONTACT: Ralph Quatrano
 Leada/Mash u Moss EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Libraries were constructed by Dr. Stavros Bashardes as part of the
 Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
 Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
 University Genome Sequencing Center for information on obtaining a
 clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
 Possible reversed clone: similarity on wrong strand
 High quality sequence stop: 332.
 Location/Qualifiers

FEATURES
 source 1..587
 /organism="Physcomitrella patens"
 /mol_type="mRNA"
 /db_xref="taxon:3218"
 /clone="PEP SOURCE ID:PPS101622"
 /dev_stage="protonemata, 7day old untreated"
 /lab_host="E.coli DH10b"
 /clone_1lb="Moss EST library PPS"
 /note="Vector: Bluescript SK-; Site 1: XhoI; Site 2:
 EcoRI; Library constructed by Stavros Bashardes and
 re-arrayed by A. Cumig & Honglin Rong. Construction of
 the cDNA library was carried out using Stratagene's 'unizap
 - cDNA synthesis kit' to ligate cDNA directionally in
 unizap XR vector arms. The vector is designed containing
 the pluescript sequence as well as the lambda DNA and
 cDNA is cloned in the EcoRI and XhoI sites in the
 pluescript sequence. The vector was then packaged using
 Gold gispackaging extracts, propagated in XL-IBase MRF
 cells and amplified. The library was excised by mass
 excision using Stratagene's Mass excision kit to infect
 SOLR cells with phagemids and ampicillin resistant
 transformants selected. Approximately 1,000,000 colonies
 were grown and recovered by using Quiagen midi prep kit. 2
 micro grams of plasmid DNA were used to transform DH10b
 cells by electroporation. Clones corresponding to abundant
 transcripts were identified by colony hybridization and
 eliminated from the library, be rearraying. This library

ORIGIN is non-directionally cloned."

Query Match 69.7%; Score 21.6; DB 12; Length 587;
 Best Local Similarity 85.7%; Pred. No. 6.7e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 AATTCAAACATTGTTCAACAGGAACC 31
 |||||
 Db 496 ACTTCAAACATTGTTCAACAGGAACC 523

RESULT 8 675 bp mRNA linear EST 24-MAY-2001
 LOCUS BE824450/c
 DEFINITION GM700023B20D7 Gm-r1070 Glycine max cDNA clone Gm-r1070-9014 3',
 mRNA sequence.

ACCESSION BE824450 GI:10256684
 VERSION BE824450.1
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max

REFERENCE 1 (bases 1 to 675)
 Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V.,
 Erpelting, J., Raph, C., Shoop, E., Partins, J., Liu, L. and Lewin, H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 Unpublished (1999)
 Other ESTs: AW568213 corresponding to Gm-r1030-4141 (5')
 CONTACT: Vodkin, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134. For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or info@genome
 systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
 Location/Qualifiers

FEATURES
 source 1..675

/organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-9014"
 /clone_1lb="Gm-r1070"
 /note="The library Gm-r1070 is a sequence-driven, rerecked
 set of 9,216 clones selected from cDNA libraries from
 various tissues and stages of development of soybean that
 represent 2,639 sequences from immature cotyledons, 1,770
 from immature seed coats, 3,938 from flowers, and 869
 from young pods. The 5' ESTs of the source clones from
 the different libraries was used to select singletons, or
 a representative of each contig, which were rerecked
 to form library Gm-r1070. The cDNA clones of the rerecked
 Gm-r1070 library were then sequenced at the 3' end. The
 contig analysis to select unique genes was performed by
 the laboratory of Ernest Retzel, Center for Computational
 Genomics and Bioinformatics, University of Minnesota,
 http://www.cbc.umn.edu/ResearchProjects/soybean/index.html
 . Rerecking was performed by Genome Systems, St. Louis,
 http://www.genomesystems.com, and 3' sequencing by the
 Keck Center for Comparative and Functional Genomics,
 University of Illinois,
 http://www.life.uiuc.edu/biotech/keck.html. Note: The

corresponding 5' EST from each clone in the Gen-rt1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

ORIGIN

Query Match 69.7%; Score 21.6; DB 10; Length 675;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 AATTCAACATTGTTCAACAGAAC 31
Db 433 AATTCAACATTGTTCAACAGAAC 406

RESULT 9
CA130450/c
LOCUS
DEFINITION SCCRT1002A10.9 RT1 Saccharum officinarum cDNA clone SCCRT1002A10
5', mRNA sequence.
ACCESSION CA130450
VERSION CA130450.1 GI:35013327
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
The libraries that made SUCBST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parvud@unicamp.br
Clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 002 row: A column: 10
Seq primer: T7 Promoter Primer.
Location/Qualifiers

1. 687
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCT1002A10"
/lab_host="DH10B"
/clone_11b="RT1"
/note="Organ: Root tips (0.3cm-long) from adult plants;
vector: pSport1; Site1: Salt; Site2: NotI; An
unidirectional cDNA library generated from [root tips
(0.3cm-long) from adult plants]. cDNA was prepared from
polyA+ mRNA using Superscript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a Sephadose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://succest.lad.ic.unicamp.br/public"

FEATURES

source

ORIGIN

Query Match 69.7%; Score 21.6; DB 13; Length 687;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CAATCAACATTGTTCAACAGAAC 30
|||||

Db 655 CAATCAACATTGTTCAACAGAAC 628

RESULT 10
AQ780940
LOCUS
DEFINITION HS 3138 B1 B06 TTC CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3138 Col=11 Row=D, genomic survey
sequence.

ACCESSION AQ780940
VERSION AQ780940.1 GI:5683900
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3138 row: D column: 11
Seq primer: T7
Class: BAC ends
High quality sequence stop: 689.
Location/Qualifiers

FEATURES

source

1. 689
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3138 Col=11 Row=D"
/sex="male"
/clone_11b="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in
E-Coli DH10B"

ORIGIN

Query Match 69.7%; Score 21.6; DB 28; Length 689;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ACAATCAACATTGTTCAACAGAAC 29
Db 407 AATTAATTAATTTGTTCAACAGAAC 434

RESULT 11
AZ202685/c
LOCUS
DEFINITION AZ202685 734 bp DNA linear GSS 31-AUG-2000
SP_0059_B1_F01_T7A Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library Strongylocentrotus purpuratus
genomic clone Plate=59 Col=1 Row=L, genomic survey sequence.

ACCESSION AZ202685
VERSION AZ202685.1 GI:8397603
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus
ORGANISM Strongylocentrotus purpuratus

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Echinoidea; Echinoidea; Echinoidea;
Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 734)
 AUTHORS Cameron, R.A., Mahlrae, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Pousette, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.
 TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
 MEDLINE 20402566
 PUBMED 10920195
 COMMENT Contact: Cameron, R.A., Davidson, E.H., Hood, L
 Division of Biology 156-29
 California Institute of Technology
 Pasadena California 91125, USA
 Tel: (626) 395-8421
 Fax: (626) 793-3047
 Email: acameron@caltech.edu
 Plate: 59 row: L column: 1
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 734.
 Location/Qualifiers
 1..734
 /organism="Strongylocentrotus purpuratus"
 /mol_type="genomic DNA"
 /db_xref="taxon:7668"
 /clone="plate=59 Col=1 Row=L"
 /clone_1ib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
 /note="Organ: Sperm; Vector: BAC3.6; BAC Clones in B-Coli DH10B"

ORIGIN
 Query Match 69.7%; Score 21.6; DB 28; Length 734;
 Best Local Similarity 85.7%; Pred. No. 6.7e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACAATTCAAACATTGTTCAACAAGAA 29
 Db 333 ACAATTCACCTTGTTCAAAGAAAGA 306

RESULT 12
 AV961084/c 283 bp mRNA linear EST 14-MAR-2002
 LOCUS AV961084 Nori Satoh unpublished cDNA library, cleavage stage embryo
 DEFINITION Clona intestinalis cDNA clone c1c114p10 5', mRNA sequence.
 ACCESSION AV961084
 VERSION AV961084.1 GI:19449383
 KEYWORDS EST.
 SOURCE Clona intestinalis
 ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cloniidae; Clona.
 1 (bases 1 to 283)
 Satoh, N., Satou, Y., Kohara, Y. and Chin, I., T.
 Expressed genes in Clona intestinalis
 Unpublished (2000)
 Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 Location/Qualifiers
 1..283
 /organism="Clona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="c1c114p10"
 /tissue_type="whole animal"
 /dev_stage="cleavage stage embryo"

FEATURES
 source

ORIGIN
 Query Match 69.0%; Score 21.4; DB 9; Length 283;
 Best Local Similarity 80.6%; Pred. No. 8.3e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TACAATTCAAACATTGTTCAACAAGAAC 31
 Db 33 TCAATTCACACCTTGTTCACACAGCAAC 3

RESULT 13
 BI161768 407 bp mRNA linear EST 09-JUL-2001
 LOCUS RE01178.3prime RE Drosophila melanogaster normalized Embryo p1C-1
 DEFINITION Drosophila melanogaster cDNA clone RE01178 3, mRNA sequence.
 ACCESSION BI161768
 VERSION BI161768.1 GI:14627574
 KEYWORDS EST.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 407)
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Friese, E., George, R., Gonzalez, M., Guerin, H., Harris, N., Li, E., Liao, G., Mista, S., Mungall, C.J., Nunoo, J., Paiele, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celinker, S. and Rubin, G.M.
 BDGP/HMI RE Drosophila EST Project
 Unpublished (2001)
 Other ESTs: RE01178.5prime
 Contact: Stapleton, M.
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
 hit genomic AE003551: arm:3L [9396216, 9682193]
 estimated-cyto:67B5-67C1: 04/11/2001
 Plate: RE.11 row: G column: 6
 High quality sequence stop: 405.
 Location/Qualifiers
 1..407
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE01178"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tora"
 /clone_1ib="RE Drosophila melanogaster normalized Embryo p1C-1"
 /note="Organ: embryo; Vector: p1C1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

ORIGIN
 Query Match 69.0%; Score 21.4; DB 12; Length 407;
 Best Local Similarity 80.6%; Pred. No. 8.1e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TACAATTCAAACATTGTTCAACAAGAAC 31
 Db 75 TACACTTAAACATTTTCAAAAGATTAAC 105

RESULT 14


```

BM195405/c      486 bp  mRNA  linear  EST 05-NOV-2002
LOCUS           BM195405 Nori Satoh unpublished cDNA library, cleaving embryo Ciona
DEFINITION      Intestinalis cDNA clone c1c1036c14 5', mRNA sequence.
ACCESSION       BM195405
VERSION         BM195405.1  GI:24609807
KEYWORDS        EST.
SOURCE          Ciona intestinalis
ORGANISM        Ciona intestinalis
                Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
                Phlebobranchia; Clonidae; Ciona.
REFERENCE       1 (bases 1 to 486)
AUTHORS         Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE           Expressed genes in Ciona intestinalis (2002c)
JOURNAL         Unpublished (2002)
COMMENT         Contact: Nori Satoh
                Department of Zoology
                Kyoto University
                Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
                Tel: 81-75-753-4081
                Fax: 81-75-705-1113
                Email: satoh@ascidian.zool.kyoto-u.ac.jp.
                Location/Qualifiers
                1..486
                /organism="Ciona intestinalis"
                /mol_type="mRNA"
                /db_xref="taxon:7719"
                /clone="c1c1036c14"
                /issue_type="whole body"
                /dev_stage="cleaving embryo"
                /clone_lib="Nori Satoh unpublished cDNA library, cleaving
                embryo"

ORIGIN
Query Match      69.0%; Score 21.4; DB 13; Length 486;
Best Local Similarity 80.6%; Pred. No. 8.1e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TACAATTCAACATGTTGTTCAACACAGCAACC 31
    | ||||| ||||| ||||| ||||| |||||
Db 36 TCCAATTCACCATGTTGTTCAACACAGCAACC 6

RESULT 15
BM209523/c      486 bp  mRNA  linear  EST 05-NOV-2002
LOCUS           BM209523 Nori Satoh unpublished cDNA library, egg Ciona
DEFINITION      Intestinalis cDNA clone c1eg061g07 5', mRNA sequence.
ACCESSION       BM209523
VERSION         BM209523.1  GI:24624137
KEYWORDS        EST.
SOURCE          Ciona intestinalis
ORGANISM        Ciona intestinalis
                Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
                Phlebobranchia; Clonidae; Ciona.
REFERENCE       1 (bases 1 to 486)
AUTHORS         Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
TITLE           Expressed genes in Ciona intestinalis (2002c)
JOURNAL         Unpublished (2002)
COMMENT         Contact: Nori Satoh
                Department of Zoology
                Kyoto University
                Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
                Tel: 81-75-753-4081
                Fax: 81-75-705-1113
                Email: satoh@ascidian.zool.kyoto-u.ac.jp.
                Location/Qualifiers
                1..486
                /organism="Ciona intestinalis"
                /mol_type="mRNA"
                /db_xref="taxon:7719"
                /clone="c1eg061g07"
                /issue_type="whole animal"

FEATURES
source

```

```

/dev_stage="egg"
/clone_lib="Nori Satoh unpublished cDNA library, egg"

ORIGIN
Query Match      69.0%; Score 21.4; DB 13; Length 486;
Best Local Similarity 80.6%; Pred. No. 8.1e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TACAATTCAACATGTTGTTCAACACAGCAACC 31
    | ||||| ||||| ||||| ||||| |||||
Db 36 TCCAATTCACCATGTTGTTCAACACAGCAACC 6

Search completed: July 20, 2004, 01:42:49
Job time : 1424.41 secs

```

U.S. DEPARTMENT OF COMMERCE

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:23:29 ; Search time 172.279 Seconds
(without alignments)
877.864 Million cell updates/sec

Title: US-09-831-272-11

Perfect score: 31

Sequence: 1 tacatcaaacattgttcaacaggaac 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/1/pubpna/PC7_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/1/pubpna/PC7US_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	21.6	69.7	91352	17	US-10-300-611-4
2	21	67.7	591	16	US-10-027-632-41058
3	21	67.7	591	13	US-10-027-632-63246
4	21	67.7	591	13	US-10-027-632-63857
5	21	67.7	591	13	US-10-027-632-179516
6	21	67.7	591	13	US-10-027-632-41058
7	21	67.7	591	16	US-10-027-632-41058
8	21	67.7	591	16	US-10-027-632-63246
9	21	67.7	591	16	US-10-027-632-63857
10	21	67.7	591	16	US-10-027-632-179516
11	21	67.7	591	16	US-10-027-632-310281
12	20.8	67.1	638	16	US-10-027-632-271776
13	20.8	67.1	638	16	US-10-027-632-271776
14	20.8	67.1	871	13	US-10-027-632-163062

15	20.8	67.1	871	13	US-10-027-632-163063	Sequence 163063,
16	20.8	67.1	871	16	US-10-027-632-163062	Sequence 163062,
17	20.8	67.1	871	16	US-10-027-632-163063	Sequence 163063,
C 18	20.4	65.8	1691139	15	US-10-067-514-1	Sequence 1, Appl1
C 19	20.4	65.8	1691139	16	US-10-419-723-1	Sequence 1, Appl1
20	20.2	65.2	342	13	US-10-335-977-4117	Sequence 4117, Ap
21	20.2	65.2	345	13	US-10-335-977-4116	Sequence 4116, Ap
22	20.2	65.2	363	13	US-10-335-977-4118	Sequence 4118, Ap
23	20.2	65.2	399	13	US-10-282-122A-41814	Sequence 41814, A
C 24	20.2	65.2	1253	13	US-10-424-599-135666	Sequence 135666,
C 25	20.2	65.2	84675	17	US-10-322-281-398	Sequence 398, App
26	20.2	65.2	2940917	13	US-10-027-632-174763	Sequence 174763,
27	20.2	65.2	2940917	16	US-10-027-632-174763	Sequence 174763,
C 28	20	64.5	406	10	US-09-918-998-3188	Sequence 3188, Ap
C 29	20	64.5	545	13	US-10-027-632-226281	Sequence 226281,
C 30	20	64.5	545	16	US-10-027-632-226281	Sequence 226281,
C 31	20	64.5	558	9	US-09-795-668-113	Sequence 113, App
C 32	20	64.5	558	9	US-09-795-668-113	Sequence 113, App
C 33	20	64.5	558	9	US-09-946-807-113	Sequence 113, App
C 34	20	64.5	612	13	US-10-027-632-271379	Sequence 271379,
C 35	20	64.5	612	13	US-10-027-632-271380	Sequence 271380,
C 36	20	64.5	612	16	US-10-027-632-271379	Sequence 271379,
C 37	20	64.5	612	16	US-10-027-632-271380	Sequence 271380,
C 38	20	64.5	747	13	US-10-027-632-271378	Sequence 271378,
C 39	20	64.5	747	16	US-10-027-632-271378	Sequence 271378,
40	20	64.5	874	16	US-10-437-038-7	Sequence 7, Appl1
41	20	64.5	874	16	US-10-437-038-8	Sequence 8, Appl1
42	20	64.5	874	16	US-10-359-091-7	Sequence 7, Appl1
43	20	64.5	874	16	US-10-359-091-8	Sequence 8, Appl1
44	20	64.5	2082	16	US-10-369-493-46177	Sequence 46177, A
45	20	64.5	5092	9	US-09-834-975-1002	Sequence 1002, Ap

ALIGNMENTS

RESULT 1
US-10-300-611-4/C
; Sequence 4, Application US/10300611
; Publication No. US20040097451A1
; GENERAL INFORMATION:
; APPLICANT: Ming-Yi Chiang
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: MODULATION OF NIDROGEN EXPRESSION
; FILE REFERENCE: PFS-0058
; CURRENT APPLICATION NUMBER: US/10/300,611
; CURRENT FILING DATE: 2002-11-19
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 4
; LENGTH: 91352
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-300-611-4

Query Match 69.7%; Score 21.6; DB 17; Length 91352;
Best Local Similarity 85.7%; Pred. No. 5.5e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACAATTCAAACATGTTCAACAGGAA 29
Db 86702 ACAATTCATTCATGTTCAACAGGAA 86675

RESULT 2
US-10-027-632-41058
; Sequence 41058, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```

; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41058
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-41058

Query Match      67.7%; Score 21; DB 13; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```

QY      2 ACAATTCAAACATGTTCAACAAGGAC 30
DB      489 AAAATTGAAAAATTGATCAGACAGGAC 517
```

```

RESULT 3
US-10-027-632-63246
; Sequence 63246, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63246
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63246
```

```

Query Match      67.7%; Score 21; DB 13; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```

QY      2 ACAATTCAAACATGTTCAACAAGGAC 30
DB      489 AAAATTGAAAAATTGATCAGACAGGAC 517
```

```

RESULT 4
US-10-027-632-63857
; Sequence 63857, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63857
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63857
```

```

Query Match      67.7%; Score 21; DB 13; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```

QY      2 ACAATTCAAACATGTTCAACAAGGAC 30
DB      489 AAAATTGAAAAATTGATCAGACAGGAC 517
```

```

RESULT 5
US-10-027-632-179516
; Sequence 179516, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179516
; LENGTH: 591
; TYPE: DNA
```

ORGANISM: Human
US-10-027-632-179516

Query Match 67.7%; Score 21; DB 13; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATTCAAACATGTTCAACAAGAAC 30
DB 489 AAAATTGAAAAATTGATCAGACAGAAC 517

RESULT 6
US-10-027-632-310281

Sequence 310281, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 310281

LENGTH: 591

TYPE: DNA

ORGANISM: Human

US-10-027-632-310281

Query Match 67.7%; Score 21; DB 13; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATTCAAACATGTTCAACAAGAAC 30
DB 489 AAAATTGAAAAATTGATCAGACAGAAC 517

RESULT 7
US-10-027-632-41058

Sequence 41058, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 41058

LENGTH: 591

TYPE: DNA

ORGANISM: Human

US-10-027-632-41058

Query Match 67.7%; Score 21; DB 16; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATTCAAACATGTTCAACAAGAAC 30
DB 489 AAAATTGAAAAATTGATCAGACAGAAC 517

RESULT 8
US-10-027-632-63246

Sequence 63246, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 63246

LENGTH: 591

TYPE: DNA

ORGANISM: Human

US-10-027-632-63246

Query Match 67.7%; Score 21; DB 16; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATTCAAACATGTTCAACAAGAAC 30
DB 489 AAAATTGAAAAATTGATCAGACAGAAC 517

RESULT 9
US-10-027-632-63857

Sequence 63857, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome

```
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 63857
LENGTH: 591
TYPE: DNA
ORGANISM: Human
US-10-027-632-63857
```

```
Query Match
Best Local Similarity 67.7%; Score 21; DB 16; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 2 ACAATTCAAACTGTTGTCACAGAGAAC 30
Db 489 AAAATTGAAAAATTGATCAGACAGAGAAC 517
```

```
RESULT 10
US-10-027-632-179516
Sequence 179516, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 179516
LENGTH: 591
TYPE: DNA
ORGANISM: Human
US-10-027-632-179516
```

```
Query Match
Best Local Similarity 67.7%; Score 21; DB 16; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 2 ACAATTCAAACTGTTGTCACAGAGAAC 30
```

```
Db 489 AAAATTGAAAAATTGATCAGACAGAGAAC 517
```

```
RESULT 11
US-10-027-632-310281
Sequence 310281, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 310281
LENGTH: 591
TYPE: DNA
ORGANISM: Human
US-10-027-632-310281
```

```
Query Match
Best Local Similarity 67.7%; Score 21; DB 16; Length 591;
Best Local Similarity 82.8%; Pred. No. 3.3e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 2 ACAATTCAAACTGTTGTCACAGAGAAC 30
Db 489 AAAATTGAAAAATTGATCAGACAGAGAAC 517
```

```
RESULT 12
US-10-027-632-271776
Sequence 271776, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 271776
```

LENGTH: 638
TYPE: DNA
ORGANISM: Human
US-10-027-632-271776

Query Match 67.1%; Score 20.8; DB 13; Length 638;
Best Local Similarity 91.7%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 TTCAAACATTGTTCAACAAGGAA 29
DB 55 TTCAAACCGTGTTCAAACAAGGAA 78

RESULT 13
US-10-027-632-271776

Sequence 271776, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 271776
LENGTH: 638
TYPE: DNA
ORGANISM: Human
US-10-027-632-271776

Query Match 67.1%; Score 20.8; DB 16; Length 638;
Best Local Similarity 91.7%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 TTCAAACATTGTTCAACAAGGAA 29
DB 55 TTCAAACCGTGTTCAAACAAGGAA 78

RESULT 14
US-10-027-632-163062

Sequence 163062, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 163062
LENGTH: 871
TYPE: DNA
ORGANISM: Human
US-10-027-632-163062

Query Match 67.1%; Score 20.8; DB 13; Length 871;
Best Local Similarity 91.7%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 TTCAAACATTGTTCAACAAGGAA 29
DB 50 TTCAAACCGTGTTCAAACAAGGAA 73

RESULT 15
US-10-027-632-163063

Sequence 163063, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 163063
LENGTH: 871
TYPE: DNA
ORGANISM: Human
US-10-027-632-163063

Query Match 67.1%; Score 20.8; DB 13; Length 871;
Best Local Similarity 91.7%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 TTCAAACATTGTTCAACAAGGAA 29
DB 50 TTCAAACCGTGTTCAAACAAGGAA 73

Search completed: July 20, 2004, 01:48:41
Job time : 175.279 secs

THE
HOLY
TRINITY
CHURCH
OF
THE
CITY
OF
NEW
YORK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:16:33 ; Search time 36.5902 Seconds

(without alignments)
470.167 Million cell updates/sec

Title: US-09-831-272-11

Perfect score: 31

Sequence: 1 tacaatcaaacactgttcaacaagaagacc 31

Scoring table: IDENTITY_NUC

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCUTS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	20	64.5	3967	4	US-09-672-749-14	Sequence 14, Appl
2	19.6	63.2	750	4	US-09-328-352-1167	Sequence 1167, Ap
3	19.6	63.2	13868	4	US-08-956-171E-173	Sequence 173, App
4	19.6	63.2	1664976	4	US-08-916-421B-1	Sequence 1, Appl
5	19.4	62.6	3396	3	US-08-974-549A-642	Sequence 642, App
6	19.4	62.6	3396	4	US-09-721-456-642	Sequence 642, App
7	19.4	62.6	58407	4	US-08-916-421B-2	Sequence 2, Appl
8	18.8	60.6	831	3	US-08-998-416-362	Sequence 362, App
9	18.8	60.6	942	4	US-08-956-171E-531	Sequence 531, App
10	18.8	60.6	3314	4	US-09-536-882A-5	Sequence 5, Appl
11	18.8	60.6	4554	4	US-08-956-171E-200	Sequence 200, App
12	18.8	60.6	10207	1	US-08-920-812-2	Sequence 2, Appl
13	18.8	60.6	10207	1	US-08-920-827-2	Sequence 2, Appl
14	18.8	60.6	10207	1	US-08-921-177-2	Sequence 2, Appl
15	18.8	60.6	10207	1	US-08-362-577C-2	Sequence 2, Appl
16	18.8	60.6	10207	2	US-08-920-828-2	Sequence 2, Appl
17	18.8	60.6	41708	4	US-09-470-512A-3	Sequence 3, Appl
18	18.8	60.6	1664976	4	US-08-916-421B-1	Sequence 1, Appl
19	18.6	60.0	9821	4	US-08-956-171E-470	Sequence 470, Appl
20	18.6	60.0	19932	2	US-08-477-451-25	Sequence 25, Appl
21	18.6	60.0	319608	4	US-09-539-333D-1	Sequence 1, Appl
22	18.6	60.0	319608	4	US-09-679-409-1	Sequence 1, Appl
23	18.4	59.4	125	1	US-07-998-003A-76	Sequence 76, Appl
24	18.4	59.4	125	1	US-08-453-274B-76	Sequence 76, Appl
25	18.4	59.4	125	1	US-08-453-695A-76	Sequence 76, Appl
26	18.4	59.4	125	1	US-08-568-161A-76	Sequence 76, Appl
27	18.4	59.4	125	2	US-08-453-702A-76	Sequence 76, Appl

28	18.4	59.4	125	3	US-09-099-639-76	Sequence 76, Appl
29	18.4	59.4	125	5	PCT-US93-12588-76	Sequence 76, Appl
30	18.4	59.4	125	5	PCT-US95-08071-76	Sequence 76, Appl
31	18.4	59.4	235	1	US-07-985-692-3	Sequence 3, Appl
32	18.4	59.4	235	1	US-08-155-331-3	Sequence 3, Appl
33	18.4	59.4	235	1	US-08-147-710-7	Sequence 7, Appl
34	18.4	59.4	235	1	US-08-321-658B-9	Sequence 9, Appl
35	18.4	59.4	235	1	US-08-384-489-6	Sequence 6, Appl
36	18.4	59.4	235	1	US-08-334-773A-3	Sequence 3, Appl
37	18.4	59.4	235	1	US-08-424-022-3	Sequence 3, Appl
38	18.4	59.4	235	1	US-08-458-090-7	Sequence 7, Appl
39	18.4	59.4	235	2	US-08-457-887-7	Sequence 7, Appl
40	18.4	59.4	235	2	US-08-424-017B-3	Sequence 3, Appl
41	18.4	59.4	235	4	US-09-904-621-7	Sequence 7, Appl
42	18.4	59.4	235	5	PCT-US93-11696-3	Sequence 3, Appl
43	18.4	59.4	348	1	US-08-468-674B-72	Sequence 72, Appl
44	18.4	59.4	348	1	US-08-780-571-72	Sequence 72, Appl
45	18.4	59.4	379	1	US-08-468-674B-74	Sequence 74, Appl

ALIGNMENTS

```
RESULT 1
US-09-672-749-14
; Sequence 14, Application US/09672749
; Patent No. 6475734
; GENERAL INFORMATION:
; APPLICANT: Liebergessel, Matthias
; APPLICANT: Falls, Patricia L
; APPLICANT: Dong, Jian G
; APPLICANT: Li, Chun Ping
; APPLICANT: Nichols, Scott E
; TITLE OF INVENTION: POLYHYDROXYALKANOATE SYNTHASE GENES
; FILE REFERENCE: 5718-120
; CURRENT APPLICATION NUMBER: US/09/672, 749
; CURRENT FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/156, 770
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 3967
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-672-749-14

Query Match      64.5%; Score 20; DB 4; Length 3967;
Best Local Similarity 82.1%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 TACAATTCAAACACTGTTCAACAACAGA 28
Db      3895 TACGATTCAAGATAGTCAACAAGAA 3922

RESULT 2
US-09-328-352-1167/c
; Sequence 1167, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328, 352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1167
; LENGTH: 750
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1167
```

Query Match 63.2%; Score 19.6; DB 4; Length 750;
Best Local Similarity 84.6%; Pred. No. 57;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATTCAACATGTTTCAACAGCAAC 30
Db 321 ACTTAACCTGTTCACACGAC 296

RESULT 3

US-08-956-171E-173/c
Sequence 173, Application US/08956171B
Patent No. 6593114

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

Gil H. Choi

Patrick S. Dillon

Craig A. Rosen

Steven C. Barash

Michael R. Pannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM: MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA: APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA: APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION: NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION: TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 173:

SEQUENCE CHARACTERISTICS:

LENGTH: 13868 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 173:

US-08-956-171E-173

Query Match 63.2%; Score 19.6; DB 4; Length 13868;
Best Local Similarity 84.6%; Pred. No. 73;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CAATTCACATGTTTCAACAGCA 28

Db 5963 CAATTCACATGTTTCTTACATGA 5938

RESULT 4

US-08-916-421B-1/c

Sequence 1, Application US/08916421B

Patent No. 6503729

GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
Patent No. 6503729

TITLE OF INVENTION: jamaaschi

FILE REFERENCE: PB275

CURRENT APPLICATION NUMBER: US/08/916,421B

CURRENT FILING DATE: 1997-08-22

PRIOR APPLICATION NUMBER: US 60/024,428

PRIOR FILING DATE: 1996-08-22

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 1664976

TYPE: DNA

ORGANISM: Methanococcus jamaaschi

FEATURE: NAME/KEY: misc_feature

LOCATION: (28222)..(28222)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (28257)..(28258)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (84773)..(84773)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (84808)..(84808)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (84812)..(84812)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (98170)..(98170)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (98159)..(98159)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (98239)..(98239)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (98266)..(98266)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (98343)..(98343)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (103998)..(103998)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (148948)..(148948)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (163385)..(163385)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (191989)..(191989)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (191995)..(191995)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (231980)..(231980)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (234187)..(234187)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (234220)..(234220)

OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature

LOCATION: (234814)..(234814)

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398) .. (309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418) .. (309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837) .. (312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993) .. (312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226) .. (319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167) .. (559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241) .. (559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992) .. (600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708) .. (622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081) .. (657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203) .. (657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435) .. (674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442) .. (682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652) .. (713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684) .. (741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455) .. (779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676) .. (779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539) .. (855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619) .. (871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830) .. (1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846) .. (1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881) .. (1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881) .. (1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988) .. (1310988)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature
LOCATION: (1313224) .. (1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473) .. (1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491) .. (1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091) .. (1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020) .. (1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912) .. (1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734) .. (1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998) .. (1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854) .. (1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 63.2%; Score 19.6; DB 4; Length 1664976;
Best Local Similarity 84.6%; Pred. No. 92;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TACATTCAAACCTGTTCACACAG 26
Db 187177 TACATTCAAACCTGTTCACACAG 187152

RESULT 5
US-08-974-549A-642/C
Sequence 642, Application US/08974549A
Patent No. 6166178
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 642:
SEQUENCE CHARACTERISTICS:
LENGTH: 3396 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..3396
OTHER INFORMATION: /note="generic" hTRT protein encoding
OTHER INFORMATION: sequence"
US-08-974-549A-642

Query Match 62.6%; Score 19.4; DB 3; Length 3396;
Best Local Similarity 79.3%; Pred. No. 77;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACAATTCAACATTGTTCACCAAGGAC 30
DB 171 ACAACCAACATTGGGCAACCAAGGCAC 143

RESULT 6
US-09-721-456-642/C
Sequence 642, Application US/09721456
Patent No. 6617110
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin B.
Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESSES:
ADDRESSES: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco

STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/721,456
FILING DATE: 22-NOV-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 642:
SEQUENCE CHARACTERISTICS:
LENGTH: 3396 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..3396
OTHER INFORMATION: /note="generic" hTRT protein encoding
SEQUENCE DESCRIPTION: SEQ ID NO: 642:
US-09-721-456-642

Query Match 62.6%; Score 19.4; DB 4; Length 3396;
Best Local Simlarity 79.3%; Pred. No. 77;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACAATTCAACATTGTTCACCAAGGAC 30
DB 171 ACAACCAACATTGGGCAACCAAGGCAC 143

RESULT 7
US-08-916-421B-2/C
Sequence 2, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Built et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ

Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 58407
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (6485)..(6485)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-2

Query Match 62.6%; Score 19.4; DB 4; Length 58407;
Best Local Similarity 79.3%; Pred. No. 97;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TACATTCAAACATTTGTTCAACAGGAA 29
DB 13255 TATGATTAGACATTTCTTCAAAAAAGTAA 13227

RESULT 8
US-08-998-416-362/C
Sequence 362, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION/DOCKET NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 362:
SEQUENCE CHARACTERISTICS:
LENGTH: 831 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: PAG1279RP
US-08-998-416-362

Query Match 60.6%; Score 18.8; DB 3; Length 831;
Best Local Similarity 74.2%; Pred. No. 12e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TACATTCAAACATTTGTTCAACAGGAAAC 31
DB 825 TAAANCAAAAGTGTCTCAATCAAGGAAC 795

RESULT 9
US-08-956-171E-531

Sequence 531, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gill H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION/DOCKET NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 531:

SEQUENCE CHARACTERISTICS:

LENGTH: 942 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 531:
US-08-956-171E-531

Query Match 60.6%; Score 18.8; DB 4; Length 942;
Best Local Similarity 76.7%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TACATTCAAACATTTGTTCAACAGGAAAC 30
DB 535 TTCAATGAACCTATGTTAAACAGGAAC 564

RESULT 10
US-09-536-882A-5/c
Sequence 5, Application US/09536882A
Patent No. 6489151
GENERAL INFORMATION:
APPLICANT: The Research Foundation of State University of New
TITLE OF INVENTION: A BIOLOGICALLY ACTIVE ALTERNATIVE FORM OF THE IKKA IKB
FILE REFERENCE: Seq. List 1-20 178-282
CURRENT APPLICATION NUMBER: US/09/536,882A
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 5
LENGTH: 3314
TYPE: DNA
ORGANISM: Mouse
US-09-536-882A-5

Query Match 60.6%; Score 18.8; DB 4; Length 3314;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACAATTCACATGTTTCAACCAAGAAC 31
DB 2924 AGACATTAAGATTGTTTAACAGTAAC 2895

RESULT 11
US-08-956-171E-200
Sequence 200, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gill H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Farnon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: P8248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1124
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:

LENGTH: 4594 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-08-956-171E-200

Query Match 60.6%; Score 18.8; DB 4; Length 4594;
Best Local Similarity 90.9%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AATTCACATGTTTCAACCA 25
DB 1150 AATTCATCATGTTTCAACCA 1171

RESULT 12
US-08-920-812-2
Sequence 2, Application US/08920812
Patent No. 5763188
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsubisa, Akio
APPLICANT: Uenata, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,812
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rln-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-24
US-08-920-812-2

Query Match 60.6%; Score 18.8; DB 1; Length 10207;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AATTCACATGTTTCAACCA 25
DB 3737 AATTCATCATGTTTCAACCA 3758

RESULT 13
US-08-920-827-2
Sequence 2, Application US/08920827
Patent No. 5770375
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotugu
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,827
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-24
US-08-920-827-2

Query Match 60.6%; Score 18.8; DB 1; Length 10207;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AATTCAACATTGTTCAACAA 25
DB 3737 AATTCATCATGTTCAACCA 3758

RESULT 14
US-08-921-177-2
Sequence 2, Application US/08921177
Patent No. 5798211
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotugu
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/921,177
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-24
US-08-921-177-2

Query Match 60.6%; Score 18.8; DB 1; Length 10207;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AATTCAACATTGTTCAACAA 25
DB 3737 AATTCATCATGTTCAACCA 3758

RESULT 15
US-08-362-577C-2
Sequence 2, Application US/08362577C
Patent No. 5807673
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotugu
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,577C
FILING DATE: 27-MAR-1995

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rln-Laure9, Ll-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-24
US-08-362-577C-2

Query Match 60.6%; Score 18.8; DB 1; Length 10207;
Best Local Similarity 90.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AATTCAACATGTTCAACAA 25
|||
Db 3737 AATTCAATCATGTTCACCA 3758

Search completed: July 20, 2004, 00:55:59
Job time : 38.5902 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM-nucleic - nucleic search, using sw model

Run on: July 20, 2004, 00:15:28 ; Search time 170.246 Seconds
(without alignments)
773.552 Million cell updates/sec

Title: US-09-831-272-11

Perfect score: 1 tacaatcaacatggttcaacaagaac 31

Sequence: 1 tacaatcaacatggttcaacaagaac 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2002s:*
6: geneseqn2003s:*
7: geneseqn2004s:*
8: geneseqn2005s:*
9: geneseqn2006s:*
10: geneseqn2007s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	AAA27974	Aaa27974 Box D str
2	21.4	69.0	2000	ACC61568	Acc61568 Gene sequ
3	21.4	69.0	4341	ABL20396	Abi20396 Drosophila
4	21.4	69.0	12010	ABL06138	Abi06138 Drosophila
5	21.2	68.4	753	AAAC33728	Aac33728 Arabidops
6	21	67.7	9507	ABN80242	Abn80242 Human che
7	20.4	65.8	541	ABV61041	Abv61041 Human pro
8	20.4	65.8	110000	ABX08336_03	Continuation (4 of
9	20.2	65.2	342	AAT67714	Aat67714 H. pylori
10	20.2	65.2	363	AAT67949	Aat67949 H. pylori
11	20.2	65.2	399	ACA53944	Aca53944 Prokaryot
12	20	64.5	184	ACH12933	Ach12933 Human sec
13	20	64.5	406	ACH15976	Ach15976 Human ada
14	20	64.5	558	AAK95314	Aak95314 Human NRG
15	20	64.5	558	AAK96807	Aak96807 Human NRG
16	20	64.5	558	ABT00084	Abt00084 Human neu
17	20	64.5	558	ABT01577	Abt01577 Human neu
18	20	64.5	658	AAAN60662	Aan60662 Sequence
19	20	64.5	814	AAAS4544	Aas4544 Human pre
20	20	64.5	844	AAAN50285	Aan50285 DNA encod
21	20	64.5	848	AAAN81387	Aan81387 Sequence
22	20	64.5	873	AAQ90324	Aaq90324 MF-alpha-
23	20	64.5	874	AAAN80450	Aan80450 Sequence

24	20	64.5	874	1	AAAN80451	Aan80451 Sequence
25	20	64.5	874	2	AAQ90325	Aaq90325 MF-alpha-
26	20	64.5	874	3	AAAS4543	Aas4543 MF1::Huma
27	20	64.5	1102	1	AAAN40048	Aan40048 Sequence
28	20	64.5	2000	7	ACC60957	Acc60957 Gene sequ
29	20	64.5	3967	4	AAAD03830	Aad03830 Saccharom
30	20	64.5	3967	4	AAAD07041	Aad07041 Saccharom
31	20	64.5	4370	9	ADCS1947	Adcs1947 Human pos
32	20	64.5	4941	9	ADCS1946	Adcs1946 Human pos
33	20	64.5	5092	4	AAAS61001	Aas61001 Human can
34	20	64.5	5575	6	AAAS61207	Aas61207 Human gen
35	20	64.5	6241	4	AAAL04604	Aal04604 Human rep
36	20	64.5	6241	4	ABL97527	Abi97527 Human tes
37	20	64.5	7525	4	ABLI7686	Abi17686 Drosophila
38	20	64.5	110000	4	AAK95240_06	Continuation (7 of
39	20	64.5	110000	4	AAK96733_06	Continuation (7 of
40	20	64.5	110000	6	ABT00010_06	Continuation (7 of
41	20	64.5	110000	6	ABT01503_06	Continuation (7 of
42	19.8	63.9	12791	4	ABLI1557	Abi11557 Drosophila
43	19.8	63.9	17282	4	ABLI1556	Abi11556 Drosophila
44	19.6	63.2	365	9	ADCS07038	Adcs07038 Novel cod
45	19.6	63.2	706	2	AAV81170	Aav81170 DNA seque

ALIGNMENTS

RESULT 1	AAA27974	standard; DNA; 31 BP.
ID	AAA27974	
XX	AAA27974;	
AC	15-AUG-2000 (first entry)	
DT		
XX		
XX	Box D strong elicitor-responsive cis-element nucleotide sequence.	
DE		
XX		
XX	Box D; elicitor-responsive cis-element; parsley; PRI promoter; ss;	
KW	chimeric promoter; pathogen infection; transgenic plant; resistance;	
KW	herbicide; local response; genetic engineering; disease resistant crop.	
OS	Petroselinum crispum.	
XX		
XX	WO200029592-A2.	
XX		
PD		
XX		
XX	25-MAY-2000.	
XX		
PF	12-NOV-1999; 99NO-BP008710.	
PR		
PR	12-NOV-1998; 98EP-00121160.	
PR	27-AUG-1999; 99EP-00116981.	
XX		
PA	(PLAC) MAX PLANCK GRS FOERDERUNG.	
XX		
PI	Kirsch C, Logemann E, Hahlbrock K, Rushton P, Somseich I;	
XX		
XX	WPI, 2000-387804/33.	
XX		
PT	Chimeric promoters mediating gene expression in plants upon pathogen	
PT	infection, useful for transgenic plant production comprises at least one	
PT	cis-acting element to direct elicitor-specific expression.	
XX		
XX	Claim 1, Page 30, 73pp; English.	
XX		
XX	This sequence represents Box D, a strong elicitor-responsive cis-element	
CC	from the Parsley PRI promoter. The present invention relates to chimeric	
CC	promoters capable of mediating local gene expression in plants upon	
CC	pathogen infection. The chimeric promoters comprise at least one cis-	
CC	element (see AAA27964-A27979) capable of directing elicitor-specific	
CC	expression, and a minimal promoter. The chimeric promoters are useful for	
CC	producing a transgenic plant which has attained resistance or improved	
CC	resistance against a pathogen. The cis-acting element, chimeric promoter,	
CC	recombinant gene encoding the chimeric promoter, vector comprising the	

CC chimeric promoter and a compound capable of activating the chimeric
CC promoter are useful for producing pathogen resistant plants, and for
CC identifying and/or producing compounds capable of conferring induced
CC resistance to a pathogen in a plant. A compound which specifically
CC activates or inhibits genes activated in a plant when attacked by a
CC pathogen is also useful as a plant protective agent or a herbicide. The
CC chimeric promoter provides rapid and local response to pathogen attack
CC but shows negligible activity in uninfected parts of the plants and
CC therefore is most suitable for the engineering of disease resistant crops
XX
SQ Sequence 31 BP; 14 A; 7 C; 3 G; 7 T; 0 U; 0 Other;

Query Match 100.0%; Score 31; DB 3; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TACAATTCACAACATGTTCCAAACAGGAAACC 31
DB 1 TACAATTCACAACATGTTCCAAACAGGAAACC 31

RESULT 2

ACC61568
ID ACC61568 standard; DNA; 2000 BP.

AC ACC61568;

DT 20-JUN-2003 (first entry)

DE Gene sequence #SEQ ID 1918.

XX Multiprotein complex; eukaryote; drug target; diagnosis; gene; ds.

OS Saccharomyces cerevisiae.

PN EP1258494-A1.

PD 20-NOV-2002.

PF 20-DEC-2001; 2001EP-00130253.

PR 15-MAY-2001; 2001EP-00111774.

PA (CELL-) CELLZONE AG.

PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;

PI Marzloch M, Schultz JD, Superti-Furga GP;

DR WPI; 2003-250078/25.

DR P-PSDB; ABR53526.

XX New isolated protein complexes useful for diagnosing a disease or
PT disorder, or as a target for an active agent of a pharmaceutical,
PT preferably a drug target in the treatment or prevention of disease or
PT disorder.

XX Disclosure; SEQ ID NO 1918; 17pp + Sequence Listing; English.

CC The invention relates to multiprotein complexes from eukaryotes. Proteins
CC of the invention and DNA sequences encoding them are given in records
CC ABR52568-ABR53903 and ACC610-ACC61944 respectively. The complexes are
CC obtainable by using a protein as a bait and isolating the set of proteins
CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM

SQ Sequence 2000 BP; 768 A; 367 C; 320 G; 545 T; 0 U; 0 Other;

Query Match 69.0%; Score 21.4; DB 7; Length 2000;
Best Local Similarity 80.6%; Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TACAATTCACAACATGTTCCAAACAGGAAACC 31
DB 1666 TACAATTCAAAATAGTTCAAAAAGAAAC 1696

RESULT 3

ABL20396/c
ID ABL20396 standard; DNA; 4341 BP.

AC ABL20396;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12661.

XX Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ds.

XX Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

XX Claim 1; SEQ ID NO 12661; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABR57737-
CC ABR72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4341 BP; 1299 A; 863 C; 857 G; 1322 T; 0 U; 0 Other;

Query Match 69.0%; Score 21.4; DB 4; Length 4341;
Best Local Similarity 80.6%; Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TACAATTCACAACATGTTCCAAACAGGAAACC 31
DB 4019 TACAATTCACAACATGTTCCAAACAGGAAAC 3989

RESULT 4

ABL06138/c
ID ABL06138 standard; cDNA; 12010 BP.

AC ABL06138;

XX

DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 12896.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacological; gene; ss.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001;
PD
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR P-PSDB; ABB62035.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
PS
XX Claim 1; SEQ ID NO 12896; 21np + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (AB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SO Sequence 12010 BP; 3743 A; 2254 C; 2323 G; 3690 T; 0 U; 0 Other;
Query Match 69.0%; Score 21.4; DB 4; Length 12010;
Best Local Similarity 80.6%; Pred. No. 1.2e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 1 TACATTCGAACATTGTTCAACAGAAC 31
Db 11688 TACACTTTAAACATTTTCAAAAAAGTAAC 11658
RESULT 5
AAC33728 ID AAC33728 standard; DNA; 753 BP.
XX
XX AAC33728;
AC
XX 17-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 4105.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX

PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 16-APR-1999; 99US-0128714P.
PR 19-APR-1999; 99US-0129845P.
PR 21-APR-1999; 99US-0130077P.
PR 23-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144864P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145182P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147453P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152353P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.

PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.4%; Score 21.2; DB 3; Length 753;
Best Local Similarity 88.5%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATTCAAACATTGTTCCAAACAGGAC 30
|||||
Db 41 ATTCAGAGATTGTTCCAAACAGGAC 66
|||||

RESULT 6
ABN80242/C
ID ABN80242 standard; DNA; 9507 BP.
XX
AC ABN80242;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 259.
XX
DE Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
XX heart disease; epilepsy; hietone deacetylation; muscular dystrophy;
XX dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
XX anticidiabetic; cytosatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007536.
XX
XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX

PX	(EPIG-)	EPIGENOMICS AG.
XX	Olek A,	Piepenbrock C, Berlin K;
XI	WI; 2002-	130908/17.
DR	Novel nucleic acid useful for diagnosis and therapy of diseases associated with development genes such as diabetes; comprises a sequence of a segment of chemically pretreated DNA of genes associated with development.	
XX	Claim 1;	SEQ ID NO 259; 27pp; English.
PS	The invention relates to a nucleic acid (I) comprising a sequence at least 18 bases in length of a segment of chemically pretreated DNA (II) of genes associated with development selected from 87 genes listed in the specification such as ACCPN, ADFN, or AFPI and comprising one of 350 sequences (ABN79984-ABN8033) or their complements. The invention is useful for the diagnosis or therapy of diseases associated with development gene, in particular disease related to homeobox containing genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes associated with congenital heart disease, epilepsy, diseases related to histone deacetylation, Curarino syndrome, diseases related with the development of the brain and limb girdle muscular dystrophy and dwarfism.	
CC	Oligomers specific to each of the genes are useful for detecting the methylation state of all CpG dinucleotides within the 350 sequences or (ii) and their complementary sequences, as primer oligonucleotides for the amplification of the 350 sequences, (iii) and/or their complements and as oligomer probes for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs). Note: The sequence data for this patent did not form part of the printed specification but is based on sequence information supplied to Derwent by the European Patent Office	
CC	Sequence 9507 BP; 2235 A; 326 C; 2497 G; 4449 T; 0 U; 0 Other;	
SQ	Query Match 67.7%; Score 21; DB 6; Length 9507; Best Local Similarity 82.8%; Pred. No. 1.7e+02; Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	
QY	2 ACAATTCAACATTTGTTCACAAGAGAAC 30 Db 1899 ACAATTCAAACATATTCAAACCATACC 1871	
DE	RESULT 7 ABV61041/c ID ABV61041 standard; CDNA; 541 BP. AC ABV61041; XX DT 13-SEP-2002 (first entry) XX DE Human prostate expression marker cDNA 61032. XX KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; phamacodgenomic marker; gene; ss. XX OS Homo sapiens. XX WN WO200160860-A2. XX PD 23-AUG-2001. XX PF 20-FEB-2001; 2001WO-US005171. XX PR 17-FEB-2000; 2000US-0183319P. PR 16-MAR-2000; 2000US-0189862P. PR 25-MAY-2000; 2000US-0207454P. PR 09-JUN-2000; 2000US-0211314P. PR 18-JUL-2000; 2000US-0219007P. PR 13-DEC-2000; 2000US-0255281P. XX VA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	

[illegible]

DT 17-JUL-1997 (first entry)
 XX H. pylori secreted or periplasmic protein ORF 80257.aa.
 XX
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
 XX
 OS Helicobacter pylori.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..342
 FT /*tag= a
 FT /note= "no stop codon given"
 XX
 XX
 PN WO9640893-A1.
 PD 19-DEC-1996.
 XX
 XX
 PF 06-JUN-1996; 96MO-US009122.
 XX
 PR 07-JUN-1995; 95US-00487032.
 PR 01-APR-1996; 96US-00630405.
 XX
 PA (ASTR) ASTR A AB.
 XX
 PI Smith D, Berglindh OT, Mellgaerd BL;
 XX
 DR WPI; 1997-052306/05.
 DR P-PSDB; AAW20571.
 XX
 XX
 PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 PT useful for vaccines to treat or prevent H. pylori infection, and to
 PT detect Helicobacter.
 XX
 PS Claim 23; Page 317; 1481pp; English.
 XX
 CC This sequence encodes a H. pylori secreted or periplasmic protein. The
 CC protein may be used in a vaccine to prevent or treat H. pylori infection
 CC or to identify H. pylori polypeptide binding compounds, useful as
 CC potential H. pylori life cycle activators or inhibitors. The genomic
 CC sequence of H. pylori (ATCC 55679) was determined from overlapping
 CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts
 CC
 XX
 SQ Sequence 342 BP; 106 A; 64 C; 44 G; 128 T; 0 U; 0 Other;
 XX
 QY Query Match 65.2%; Score 20.2; DB 2; Length 342;
 Best Local Similarity 88.0%; Pred. No. 2.9e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 XX
 Db 1 TACAATTCAACATTGTTCAACAA 25
 38 TACAATTCAACATTGTTCAACAA 62
 XX
 RESULT 10
 AAT67949
 ID AAT67949 standard; DNA; 363 BP.
 XX
 AC AAT67949;
 XX
 DT 15-JUL-1997 (first entry)
 XX
 XX H. pylori secreted or periplasmic protein ORF 05ae20220orf50.
 DE
 XX

KW Secretion; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
 KW periplasm; ds.
 XX
 OS Helicobacter pylori.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..363
 FT /*tag= a
 FT /note= "no stop codon given"
 XX
 XX
 PN WO9640893-A1.
 PD 19-DEC-1996.
 XX
 XX
 PF 06-JUN-1996; 96MO-US009122.
 XX
 PR 07-JUN-1995; 95US-00487032.
 PR 01-APR-1996; 96US-00630405.
 XX
 PA (ASTR) ASTR A AB.
 XX
 PI Smith D, Berglindh OT, Mellgaerd BL;
 XX
 DR WPI; 1997-052306/05.
 DR P-PSDB; AAW20696.
 XX
 XX
 PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 PT useful for vaccines to treat or prevent H. pylori infection, and to
 PT detect Helicobacter.
 XX
 PS Claim 23; Page 806; 1481pp; English.
 XX
 CC The present sequence encodes a Helicobacter pylori protein which is
 CC likely to be secreted or periplasmic. The protein may be used in a
 CC vaccine to prevent or treat H. pylori infection or to identify H. pylori
 CC polypeptide binding compounds, useful as potential H. pylori life cycle
 CC activators or inhibitors. The genomic sequence of H. pylori (ATCC 55679)
 CC was determined from overlapping contigs generated by mechanically
 CC shearing the bacterial DNA. The sequences were analysed for ORF of at
 CC least 180 nucleotides, and the predicted coding regions defined by
 CC computer evaluation. To identify likely H. pylori antigens for vaccine
 CC development, the amino acid sequences predicted from various ORF were
 CC analysed for significant homology to other known or exported membrane
 CC proteins. Having identified and determined the sequences of interest,
 CC particular regions can be isolated from H. pylori by PCR amplification
 CC for recombinant polypeptide production, e.g. in E. coli hosts
 CC
 XX
 SQ Sequence 363 BP; 116 A; 69 C; 44 G; 134 T; 0 U; 0 Other;
 XX
 QY Query Match 65.2%; Score 20.2; DB 2; Length 363;
 Best Local Similarity 88.0%; Pred. No. 2.9e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 XX
 Db 1 TACAATTCAACATTGTTCAACAA 25
 59 TACAATTCAACATTGTTCAACAA 83
 XX
 RESULT 11
 ACAS3944
 ID ACAS3944 standard; DNA; 399 BP.
 XX
 AC ACAS3944;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 XX Prokaryotic essential gene #35601.
 DE
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX

OS Versinia pectis.
 XX PN W0200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002MO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX PA Wang L, Zamudio C, Malone C, Haeselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
 XX P-PSDB; ABUS0074.
 DR MPI; 2003-029926/02.
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 1; SEQ ID NO 41814; 1766bp; English.
 XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 399 BP; 102 A; 98 C; 95 G; 104 T; 0 U; 0 Other;

Query Match 65.2%; Score 20.2; DB 7; Length 399;
 Best Local Similarity 88.0%; Pred. No. 2.9e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TCAACATGTTTCAACAGGAAAC 31
 DB 363 TCATACATTGTTCCACAGGAAAC 387

RESULT 12
 AAC12933/c

ID AAC12933 standard; cDNA; 184 BP.
 XX AC AAC12933;
 XX DT 06-OCT-2000 (first entry)
 XX DE Human secreted protein 5' EST, SEQ ID NO: 17008.
 XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX OS Homo sapiens.
 XX PN EP1033401-A2.
 XX PD 06-SEP-2000.
 XX PF 21-FEB-2000; 2000EP-00200610.
 XX PR 26-FEB-1999; 99US-0122487P.
 XX (GEST) GENSET.
 XX PA Dumas Milne Edwards J, Duclert A, Giordano J;
 PI MPI; 2000-500381/45.
 DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX Claim 1; SEQ ID NO 17008; 71bp + Sequence Listing; English.
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors
 XX
 SQ Sequence 184 BP; 53 A; 36 C; 44 G; 51 T; 0 U; 0 Other;

Query Match 64.5%; Score 20; DB 3; Length 184;
 Best Local Similarity 82.1%; Pred. No. 3.3e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATTCAACATGTTTCAACAGGAA 29
 DB 68 ACAATTCAACATGTTTCAACAGGAA 41

RESULT 13
 ACH15976/c
 ID ACH15976 standard; cDNA; 406 BP.
 XX AC ACH15976;
 XX DT 13-OCT-2003 (first entry)
 XX DE Human adult heart cDNA #290.
 XX KW Human; ss; sequencing by hybridisation; SHH; expressed sequence tag; EST;
 KW genome mapping; biodiversity; genetic disorder.
 XX OS Homo sapiens.

```
XX US2003073623-A1.
PN
XX
XX 17-APR-2003.
PD
XX
XX 30-JUL-2001; 2001US-00918995.
PF
XX
XX 30-JUL-2001; 2001US-00918995.
PR
XX
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
DR
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 3188; 44pp; English.
PS
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a useful frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
CC
XX
SQ Sequence 406 BP; 148 A; 70 C; 88 G; 100 T; 0 U; 0 Other;
Query Match 64.5%; Score 20; DB 8; Length 406;
Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 2 ACAATTCAACATGTGTTCAACAAGGA 29
DB 388 ACTGTTCAACCTGTGTTCAACAAGGA 361
```

```
RESULT 14
ID AAK95314/c
XX AAK95314 standard; DNA; 558 BP.
AC
XX AAK95314;
XX
XX 17-DEC-2001 (first entry)
DT
XX
XX Human NRGIAG1 exon single nucleotide polymorphism E588B.
DE
XX
XX Human; neuregulin-1 associated gene 1; NRGIAG1; Schizophrenia gene;
KM gene therapy; single nucleotide polymorphism; SNP; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200164876-A2.
PN
XX
XX 07-SEP-2001.
PD
```

```
XX 28-FEB-2001; 2001WO-US006376.
PF
XX
XX 28-FEB-2000; 2000US-00515715.
PR
XX
XX (DECO-) DECODE GENETICS EHF.
PA
XX
XX Stefansson H, Steinchorodottir V, Gulcher JR;
PI
XX
XX WPI; 2001-550179/61.
DR
XX
XX Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for
PT preventing diagnosing and treating schizophrenia.
PT
XX
XX Disclosure; Page 507; 750pp; English.
PS
XX
```

This sequence represents a single nucleotide polymorphism (SNP) of the human neuregulin-1 associated gene 1 (NRGIAG1) of the invention. The NRGIAG1 gene is also referred to as the human Schizophrenia gene. The invention also relates to fragments or variants of the gene and the NRGIAG1 polypeptides they encode. The NRGIAG1 nucleic acids and polypeptides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate NRGIAG1 expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of NRGIAG1 by expressing inactive proteins or to supplement the patients own production of NRGIAG1. Additionally, the gene may be used to produce NRGIAG1 polypeptides, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The gene may also be used as DNA probes and primers in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The NRGIAG1 polypeptides may also be used as antigens in the production of antibodies against NRGIAG1 and in assays to identify modulators of NRGIAG1 expression and activity. Anti-NRGIAG1 antibodies and antagonists may also be used to down regulate expression and activity. Anti-NRGIAG1 antibodies may also be used as diagnostic agents for detecting the presence of NRGIAG1 polypeptides in samples. NRGIAG1 is associated with schizophrenia which may be prevented, diagnosed and/or treated by the above methods

```
SQ Sequence 558 BP; 185 A; 81 C; 104 G; 187 T; 0 U; 1 Other;
Query Match 64.5%; Score 20; DB 4; Length 558;
Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 2 ACAATTCAACATGTGTTCAACAAGGA 29
DB 175 ACTGTTCAACCATGTTCAACAAGGCA 148
```

```
RESULT 15
ID AAK96807/c
XX AAK96807 standard; DNA; 558 BP.
AC
XX AAK96807;
XX
XX 17-DEC-2001 (first entry)
DT
XX
XX Human NRGIAG1 exon single nucleotide polymorphism E588B.
DE
XX
XX Human; neuregulin 1 gene; schizophrenia; gene therapy; SNP;
KM single nucleotide polymorphism; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200164877-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 28-FEB-2001; 2001WO-US006377.
PF
XX
```


PR 28-FEB-2000; 2000US-00515716.

PA (DECO-) DECODE GENETICS EHF.

PI Stefnasson H, Steinhorsdottir V, Gulcher JR;

DR WPI; 2001-514841/56.

PT Neuregulin 1 nucleic acids and proteins useful for diagnosing preventing
PT and treating schizophrenia.

PT and treating schizophrenia.

PS Disclosure; Page 92; 756pp; English.

This sequence represents a single nucleotide polymorphism (SNP) from the human neuroregulin 1 gene of the invention. The invention also relates to fragments or variants of the neuroregulin 1 gene. The gene and its proteins may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate neuroregulin 1 expression, such as schizophrenia. For example they may be used to treat disorders associated with decreased neuroregulin 1 expression by rectifying mutations or deletions in a patient's genome that affect the activity of neuroregulin 1 by expressing inactive proteins or to supplement the patients own production of polypeptides. Additionally, the gene may be used to produce the neuroregulin 1 protein, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The gene and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The protein may also be used as antigens in the production of antibodies against neuroregulin 1 and in assays to identify modulators of neuroregulin 1 expression and activity. The antibodies and antagonists may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of neuroregulin 1 in samples

SQ Sequence 558 BP; 185 A; 81 C; 104 G; 187 T; 0 U; 1 Other;

Query Match	Score	DB	Length
64.58;	20;	4;	558;

Best Local Similarity 83.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0;
Gaps 0

QY	2	ACAAATTCAAACATGTTCAACCAAGGAA	29
Db	175	ACTGTTCAACCATGTTCAACCAAGGCA	148

Search completed: July 20, 2004, 00:29:14
Job time : 174.246 secs

This Page Blank (uspo:q)

TITLE Chinese promoters capable of mediating gene expression in plants upon pathogen infection and uses thereof
JOURNAL Patent: WO 0029592-A 2 25-MAY-2000;
MAX PLANCK GESAMTSCHAFT (DE) ; LOGEMANN ELKE (DE) ; SOMSICH IMRE (DE) ; HAHNBROCK KLAUS (DE) ; KIRSCH CHRISTOPH (DE) ; RUSHTON PAUL (DE)

FEATURES
source 1..30
/organism="Petroselinum crispum"
/mol_type="unassigned DNA"
/db_xref="taxon:4043"

ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTAACATTCGCC 30
1 CACACTTAATTGACCGAGTAACATTCGCC 30

Db PCPR115 930 bp DNA linear PLN 30-MAY-1989
DEFINITION Parsley PCPR1-1 gene for pathogenesis-related protein type A 5'
UTR
X12572
X12572.1 GI:20455
pathogenesis-related protein.
KEYWORDS Petroselinum crispum (parsley)
SOURCE Petroselinum crispum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Apiales; Apiaceae; Apioidae; apioid superclade; Apium clade; Petroselinum.
1 (bases 1 to 930)
Somsich,I.B., Schmelzer E., Kwallack,P. and Hahlbrock,K. Gene structure and in situ transcript localization of pathogenesis-related protein 1 in parsley Mol. Gen. Genet. 213 (1), 93-98 (1988)
MEDLINE 89127150
PUBMED 3221838
COMMENT see x15085 for PR1-1 gene 3' UTR; see x12573 and x12574 for PR1-3 and PR1-2 cDNA seqs;
PR1-1 gene is organized into two exons (274 and 439 bp in size), interrupted by a 164 bp intron.
location/Qualifiers
1..930
/organism="Petroselinum crispum"
/mol_type="genomic DNA"
/db_xref="taxon:4043"
/clone="lambda psi1"
/clone_lib="lambda EMBL4."
841..>930
/note="PR1-1 transcript"
841..>930
/note="PR1-1 mRNA exon 1"
927..>930
/note="unamed protein product; Protein sequence is in conflict with the conceptual translation; PR1-1"
/codon_start=1
/protein_id="CAA31084.1"
/db_xref="GI:4379347"
/translation="Y"

ORIGIN
Query Match 100.0%; Score 30; DB 8; Length 930;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTAACATTCGCC 30
1 CACACTTAATTGACCGAGTAACATTCGCC 30

Db 589 CACACTTAATTGACCGAGTAACATTCGCC 618

RESULT 3
LOCUS PCU48862 2011 bp DNA linear PLN 12-FEB-1997
DEFINITION Petroselinum crispum pathogenesis-related protein (prl-1) gene, complete cds.
ACCESSION U48862
VERSION U48862
KEYWORDS U48862.1 GI:1840040
SOURCE
Petroselinum crispum (parsley)
Petroselinum crispum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Apiales; Apiaceae; Apioidae; apioid superclade; Apium clade; Petroselinum.
1 (bases 1 to 2011)
Rushton,P.J., Torres,J.T., Parniske,M., Werner,P., Hahlbrock,K. and Somsich,I.B. Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes EMBO J. 15 (20), 5690-5700 (1996)
MEDLINE 97051827
PUBMED 8896462
COMMENT 2 (bases 1 to 2011)
Somsich,I.B.
Submitted (09-FEB-1996) Imre F. Somsich, Dept. of Biochemistry, Max-Planck-Institut fuer Zuechtungsforshung, Carl-von-Linne Weg 10, Koeln, 50829, Germany
location/Qualifiers
1..2011
/organism="Petroselinum crispum"
/mol_type="genomic DNA"
/db_xref="taxon:4043"
113..1727
/gene="prl-1"
113..134
/gene="prl-1"
/note="elicitor-response element 2"
589..616
/gene="prl-1"
/note="elicitor-response element 1"
695..716
/gene="prl-1"
/note="constitutive in vivo DNA footprint region"
739..743
/gene="prl-1"
773..777
/gene="prl-1"
814..820
/gene="prl-1"
join(841..1114,1279..1761)
841..1761
5'UTR
/gene="prl-1"
join(931..1114,1279..1562)
/gene="prl-1"
/function="unknown"
/note="PR1-1; belongs to the PR-10 class of intracellular PR proteins"
/codon_start=1
/product="pathogenesis-related protein 1"
/protein_id="AA847234.1"
/db_xref="GI:1840041"
/translation="MGVQKSEVETTSVSAEKLFGALCLIDITLLPQVLPGAIKSSET
LEGDGAVKLVHLGDASPEFTMKQVDADIKATPTYSGLIIDGDIIGFESINNH
ETAVNADGGCTVSTIIIFNTKGAIVAEENIKFANDQNLTFKAVEAYLIAN"

ORIGIN
3'UTR
polyA_signal
/gene="prl-1"

Query Match 100.0%; Score 30; DB 8; Length 2011;
 Best Local Similarity 100.0%; Pred. No. 0.0021;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTAACATTGCGC 30
 |||
 Db 589 CACACTTAATTGACCGAGTAACATTGCGC 618

RESULT 4
 PCU48863 1312 bp DNA linear PLN 12-FEB-1997
 LOCUS Petroselinum crispum pathogenesis-related protein (prl-2) gene,
 DEFINITION promoter region and most of exon 1.
 ACCESSION U48863 GI:1840042
 VERSION U48863
 KEYWORDS Petroselinum crispum (parsley)
 SOURCE Petroselinum crispum
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asteridae; campanulids; Apiales; Apiaceae; Apioideae; apioide
 superclade; Apium clade; Petroselinum.
 1 (bases 1 to 1312)
 Rushon, P.J., Torres, J.T., Parniske, M., Wernert, P., Hahlbrock, K.
 and Somsich, I.E.
 Interaction of elicitor-induced DNA-binding proteins with elicitor
 response elements in the promoters of parsley prl genes
 EMBO J. 15 (20), 5690-5700 (1996)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 8896462
 2 (bases 1 to 1312)
 Somsich, I.E.
 Direct Submission
 Submitted (09-FEB-1996) Imre E. Somsich, Dept. of Biochemistry,
 Max-Planck-Institut fuer Zuechtungsforchung, Carl-von-Linne Weg
 10, Koeln, 50829, Germany
 Location/Qualifiers
 1..1312
 /organism="Petroselinum crispum"
 /mol_type="genomic DNA"
 /db_xref="taxon:4043"
 1..1312
 /gene="prl-2"
 1..1082
 /gene="prl-2"
 670..703
 /gene="prl-2"
 /note="elicitor-response element 3"
 814..858
 /gene="prl-2"
 /note="elicitor-response element 1"
 929..950
 /gene="prl-2"
 /note="constitutive in vivo footprinted region"
 969..973
 /gene="prl-2"
 1003..1007
 /gene="prl-2"
 1056..1062
 /gene="prl-2"
 1083..1172
 /gene="prl-2"
 1083..1172
 /gene="prl-2"
 1173..1312
 /gene="prl-2"
 /function="unknown"
 /note="prl-2; belongs to the PR-10 class of intracellular
 PR proteins"
 /codon_start=1
 /product="pathogenesis-related protein 1-2"

/protein_id="AAB47235.1"
 /db_xref="GI:1840043"
 /translation="MGVQKSEVETTSVSAAEKLFGKLCIDITPLPQVLPAIKSSET
 LFG"

ORIGIN
 Query Match 94.7%; Score 28.4; DB 8; Length 1312;
 Best Local Similarity 96.7%; Pred. No. 0.012;
 Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTAACATTGCGC 30
 |||
 Db 816 CACACTTAATTGACCGAGTAACATTGACC 845

RESULT 5
 AL670670/c 91811 bp DNA linear ROD 23-MAY-2002
 LOCUS Mouse DNA sequence from clone RP23-23407 on chromosome X, complete
 DEFINITION sequence.
 ACCESSION AL670670 GI:21213385
 VERSION AL670670.7
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 91811)
 Direct Submission
 Submitted (23-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On May 25, 2002 this sequence version replaced gi:21065347.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-23407 is
 from the RPI-23 Mouse PAC library
 constructed by the group of Pieter de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pBAC3.6.
 Location/Qualifiers
 1..91811
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-23407"
 /clone_11b="RPI-23"

ORIGIN
 Query Match 68.0%; Score 20.4; DB 10; Length 91811;
 Best Local Similarity 80.0%; Pred. No. 92;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTAACATTGCGC 30
 |||
 Db 37196 CACACTTAATTGACCGAGTAACATTGCGC 37167

RESULT 6
AL513220/c 94638 bp DNA linear PRI 17-APR-2002
LOCUS Human DNA sequence from clone RP11-566C13 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL513220
VERSION AL513220.9 GI:20196551
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Matthews, N.
TITLE Direct Submission
JOURNAL Submitted (17-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Apr 18, 2002 this sequence version replaced gi:19699427.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-566C13 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES
source
1..94638
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-566C13"
/clone_1lb="RPCT-11.2"

ORIGIN
Query Match 68.0%; Score 20.4; DB 9; Length 94638;
Best Local Similarity 80.0%; Pred. No. 92;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
1 CACACTTAATTACGAGTAACATTCGCC 30
Db 1781 CACACTTGATTTCACGAGTAACATTCAGAC 1752

RESULT 7
HS423B22/c 112224 bp DNA linear PRI 04-APR-2001
LOCUS Human DNA sequence from clone RP3-423B22 on chromosome 1p33-35.3
DEFINITION Contains part of a gene for a novel protein similar to KIAA0323, KIAA0615 and C.elegans C30F12.1, a ferritin H pseudogene, the gene for RPS27 (ribosomal protein S27 (metallopanstimulin 1)), the gene

for axonemal dynein light chain (hp28), the gene for a novel nuclear guanosine 5'-triphosphate binding protein autoantigen, STSs, GSSs and a CpG island, complete sequence.
AL034379
AL034379.8 GI:5918013
KEYWORDS HTG; autoantigen; Cpg island; dynein; ferritin; hp28; KIAA0323; KIAA0615; metallopanstimulin; ribosomal protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 112224)
REFERENCE
AUTHORS Pearce, A.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Sep 22, 1999 this sequence version replaced gi:5870406.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP3-423B22 is from the library RPCT-3 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-423B22.

FEATURES
source
1..112224
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p33-35.3"
/clone="RP3-423B22"
/clone_1lb="RPCT-3"
885..1249
/note="match: GSS: Em:AQ188117"
complement(1836..1220)
/note="match: GSS: Em:AQ124662"
join(3734..3873,5045..5279,5878..5984)
/gene="DU423B22.1"
join(3734..3873,5045..5279,5878..5984)
/gene="DU423B22.1"
/note="match: ESTs: Em:A1832215
match: proteins: Tr:O15037 Tr:O75113 Tr:O95530 Wp:CE16881"
/codon_start=2
/evidence="not_experimental"
/product="DU423B22.1 (novel protein similar to KIAA0323, KIAA0615 and C.elegans C30F12.1)"
/protein_id="CAC36083.1"
/db_xref="GI:13559173"
/db_xref="SPTREMBL:Q9B029"
/translation="HGKKEVFSRCRITLAVNMFLEKGTDTITVPVSRKQPRPDV ITDHIILREKKKILVTPSRVYGKRRVCYDRIIVKLAIESDGIIVSNDYRDLQ

```

GERQEWKRFIEERLWYSFVNDKFMPPDDBLGRHGPSLDNPLKPKLTLEHRKQPCPY
"
repeat_region
  16005..16034
  /note="15 copies 2 mer aa 86% conserved"
misc_feature
  16038..16217
  /note="match: STS: Em:G25594 Em:G27659"
misc_feature
  16758..17140
  /note="match: STS: Em:G38585"
misc_feature
  16768..17099
  /note="match: STS: Em:G29990"
  complement(16958..17522)
misc_feature
  17051..17124
  /note="match: GSS: Em:AQ046586"
repeat_region
  /note="37 copies 2 mer aa 71% conserved"
misc_feature
  complement(17893..18413)
  /note="match: GSS: Em:AQ635919"
misc_feature
  complement(17988..18405)
  /note="match: GSS: Em:AQ220399"
  complement(21756..22258)
misc_feature
  /note="match: GSS: Em:AQ804436"
  22259..22501
misc_feature
  /note="match: GSS: Em:AQ312816"
  22259..22496
misc_feature
  /note="match: GSS: Em:AQ277563"
  22272..22698
misc_feature
  /note="match: GSS: Em:AQ880377"
  25551..25607
repeat_region
  /note="3 copies 19 mer 98% conserved"
  25981..26022
repeat_region
  /note="21 copies 2 mer tt 78% conserved"
  28245..28270
  /note="13 copies 2 mer ac 100% conserved"
  complement(28745..29200)
misc_feature
  /note="match: GSS: Em:AQ279885"
  37708..37903
repeat_region
  /note="98 copies 2 mer cc 59% conserved"
  37745..38622
misc_feature
  /note="CpG island"
  /evidence=not_experimental
  39158..39335
repeat_region
  /note="LIMB3 repeat: matches 1519..1695 of consensus"
  40696..40874
repeat_region
  /note="MER51-internal repeat: matches 2762..2946 of consensus"
  40913..41325
repeat_region
  /note="MER51-internal repeat: matches 2340..2762 of consensus"
  41425..41512
repeat_region
  /note="MSTB-internal repeat: matches 53..143 of consensus"
  41513..42266
repeat_region
  /note="PTRS repeat: matches 743..1476 of consensus"
  43004..43054
repeat_region
  /note="MSTB-internal repeat: matches 6..59 of consensus"
  44443..44578
repeat_region
  /note="LIMB3 repeat: matches 1234..1410 of consensus"
  48476..48626
repeat_region
  /note="MER83-internal repeat: matches 1652..1790 of consensus"
  49239..49303
repeat_region
  /note="MER83-internal repeat: matches 1790..1857 of consensus"
  49659..49763
repeat_region
  /note="MER83-internal repeat: matches 2718..2814 of consensus"
  complement(52297..52760)
  /note="match: GSS: Em:AQ474424"
  complement(57547..57900)
misc_feature
  /note="match: GSS: Em:AQ296347"
  complement(59824..60367)
  /note="match: GSS: Em:AQ547506"
  60547..60754
  /note="match: GSS: Em:AQ473963"
  60701..60797

```

```

gene
  /note="match: STS: Em:HSPE43A11 Em:HSPE49C01"
  complement(join(61207..61456,63601..63972))
  /gene="d423B22.2"
  complement(join(61207..61456,63601..63972))
  /gene="d423B22.2"
  /note="match: proteins: Wp:CE08529"
  /codon_start=2
  /evidence=not_experimental
  /product="d423B22.2 (novel protein similar to C.elegans protein CE08529)"
  /protein_id="CAC36084.1"
  /db_xref="GI:13559174"
  /db_xref="SPTREMBL:O9BOZ8"
  /translation="RRERQRNDVGCGSESELYPRPGNNKEKVPKPKSPRLS
  GLEDLNTFRGVIVKYSEPPPEARLPKKRWLYPKKNDLVPMVTHRQSAVLGRHR
  RADLPIDHPSCKQHAVPQYRLVETRGVGRVKKVYIDLSGNGTFLNKRRIE
  PORYVEKKEQVLYKFGFSREYVYLHESDTEIDRKDEDEDEEEEEE"
  61164..65853
repeat_region
  /note="MER44C repeat: matches 1..728 of consensus"
  68077..68389
misc_feature
  /note="match: GSS: Em:AQ839815"
  68207..68745
gene
  /gene="d423B22.3"
  /pseudo
  68207..68745
  /gene="d423B22.3"
  /note="d423B22.3 (ferritin H pseudogene)"
  match: CDNA8: Em:L20941 Em:M97164 Em:E08663 Em:D86627
  Em:M12937 Em:M63912 Em:M99692 Em:DI5071 Em:J04755
  Em:M11146 Em:S77386 Em:U54800 Em:X00318 Em:AB003093
  match: BSTS: Em:A1816415 Em:A1346347 Em:A1460154
  Em:AA632163 Em:A1775128 Em:A1092944 Em:AA057837
  Em:A1929679 Em:A1935207 Em:AA642200 Em:A145942
  Em:AA010815 Em:A1346840 Em:A1582308 Em:A170636
  Em:A1246580 Em:A131039 Em:AA604610 Em:A1079884 Em:757843
  Em:AA056594 Em:A1830614 Em:N78958 Em:A1878024 Em:A1749039
  Em:A1864456 Em:A161102 Em:C18478 Em:A1878269 Em:AA165061
  Em:AM021110 Em:AA963209 Em:AA725087 Em:AA157187
  Em:AA083658 Em:D51220 Em:AA186730 Em:A1950656 Em:AA122004
  Em:AA565350 Em:AA084238 Em:A1247121 Em:AA086011 Em:AA456170
  Em:A1075450 Em:AA657366 Em:AA151760 Em:A1815565
  Em:A1785093 Em:A1929672 Em:N70862 Em:A1939639 Em:AA102093
  Em:AA112615 Em:AA701506 Em:W31508 Em:AA831015 Em:AA044647
  Em:AA602089 Em:A1929641 Em:AA052988 Em:AA482667 Em:299431
  Em:A1831973 Em:C17972 Em:AA829459 Em:C18189 Em:AA225120
  Em:AA296646 Em:AA662010 Em:AA191015 Em:AA303930 Em:R67061
  Em:AA1889320 Em:A173411 Em:AA371395 Em:AA918127 Em:N89764
  Em:AA057849 Em:D54420 Em:AA808377 Em:A1876587 Em:AA132507

```

Query Match

Best Local Similarity 68.0%; Score 20.4; DB 9; Length 112224;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTAACATTCGCC 30
DB 112005 CACACTTGATTATCCAGTACATTAGAC 111976

RESULT 8

ALJ355880/c 153023 bp DNA linear HTG 10-UTL-2001
LOCUS Homo sapiens chromosome 1 clone RP1-195A8, 4 unordered pieces.
DEFINITION ALJ355880 AC027731
ACCESSION ALJ355880.7 GI:11876042
VERSION HTG; HTGS PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Wallis J.
AUTHORS Direct Submission
TITLE

JOURNAL

COMMENT

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On or before May 15, 2001 this sequence version replaced
gi:8077031, gi:11544491.
Draft Sequence produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA195A8

----- Summary Statistics
Assembly program: XGAP4, version 4.5
Sequencing vector: M13; M77815; 34% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 150796 bases at least Q40
Consensus quality: 151533 bases at least Q30
Consensus quality: 151958 bases at least Q20
Insert size: 152723; sum-of-contigs
Insert size: 155229; 2.8% error; agarose-fp
Quality coverage: 9.72x in Q20 bases; sum-of-contigs Quality
coverage: 9.97x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preselected.

* 1 144938: contig of 144938 bp in length
* 144939 145038: gap of 100 bp
* 145039 147529: contig of 2491 bp in length
* 147530 147629: gap of 100 bp
* 147630 149636: contig of 2007 bp in length
* 149637 149736: gap of 100 bp
* 149737 153023: contig of 3287 bp in length.

FEATURES

source

1.153023
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-195A8"
/clone_11b="RPCT-11.1"
1.144938
/note="assembly_fragment:00424
clone_end:17
vector_side:left"
145039..147529
/note="assembly_fragment:00960"
147630..149636
/note="assembly_fragment:01053"
149737..153023
/note="assembly_fragment:03302"

ORIGIN

Query Match 68.0%; Score 20.4; DB 2; Length 153023;
Best Local Similarity 80.0%; Pred. No. 92;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACGAGTACATTCGCC 30
DB 141556 CACACTTAATTGACGAGTACATTCGCC 141527

RESULT 9

BX571867

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BX571867 336609 bp DNA linear BCT 26-SEP-2003
Phototrichabius luminescens subsp. laumondii TT01 complete genome;
segment 9/17.

BX571867 BX470251
BX571867.1 GI:36785705
complete genome.
Phototrichabius luminescens subsp. laumondii TT01
Phototrichabius luminescens subsp. laumondii TT01
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Phototrichabius.

1
Duchaud,E., Rusniok,C., Frangeul,L., Buchrieser,C., Taourit,S.,
Bocs,S., Bouraux-Eude,C., Chandler,M., Dassa,B., Derose,R.,
Dzerelle,S., Freysinet,G., Gaudriault,S., Givaudan,A., Glaeser,P.,
Medigue,C., Lanolis,A., Powell,K., Siguier,P., Wingate,V.,
Zouine,M., Boemare,N., Danchin,A. and Kunst,F.
Complete genome sequence of the entomopathogenic bacterium
Phototrichabius luminescens
Nat. Biotechnol. 11 (1) (2003) In press

2
Duchaud,E., Frangeul,L., Rusniok,C. and Kunst,F.
Direct Submission
Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25
rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:
lfrangeul@pasteur.fr, fkunst@pasteur.fr

location/Qualifiers

1..336609

/organism="Phototrichabius luminescens subsp. laumondii TT01"

/mol_type="genomic DNA"

/strain="TT01"

/db_xref="taxon:243265"

complement(425..1409)

/locus_tag="plu2374"

complement(425..1396)

/locus_tag="plu2374"

/function="unknown"

/note="unnamed protein product; Highly similar to putative
cytochrome oxidase, yfeH protein of Escherichia coli"

/codon_start=1

/transl_table=11

/protein_id="CAE14684.1"

/db_xref="GI:36785706"

/translation="MTWMOKLKIDSFILMTLVVVVIAFPCEGSIKWPORLTAAI
ALFFMHGAKLSROALITAGMAHRLHIVPSTFVLPILIGLNLVLPIMTSVVM
GPIYICALPATYQSHIAFTSYAGCNVAAICSAASSTIGPISIIIGLMTHTGCG
SADTIEAIKAITIQLMVPFIIGHSRPLIARWVEGKLLMTDRSSITLVVYAFSE
AVVEGIVKHKIDGSLITIGVCCVLAVLVINTYIARMLGFKKDEITIVFCGSKS
LANGVPMANVLFPATVGMPLPLMIFHQIOLMCAVALORVAKLTK"

complement(1404..1409)

/locus_tag="plu2374"

complement(1605..2653)

/locus_tag="plu2375"

complement(1605..2642)

/locus_tag="plu2375"

/function="unknown"

/note="unnamed protein product; Similar to hypothetical
oxidoreductase YdgJ of Escherichia coli"

/codon_start=1

/transl_table=11

/protein_id="CAE14688.1"

/db_xref="GI:36785707"

/translation="MRRLKVGIVGYASTFPAHPLIAGTAGVTLAIISSDAEKV
SDMNPVIVSRPREDIFNDPDIIDLIYIPFNTHTYPLAKALAAAGHVVDKPTTVE
QAESLKQQAERADLLSVFNHRMWSGFLTKSLIKERLGLTKFASHFVRITIR
QMRERAAQSGSIGIWDLPHLIDQAIQLFGKQSLTADLGMIRPQAEVADVFEVILIR
PDIKVILHATTLAAESAVYTLHGAGSYVYKGLDPOEERLKAGRRPDAHWGYDIRD
GIVTLSHGEBLVTKHVPFLPGNYSAYVAIIDAIMLGKPNVPTANEALIQIMLIRAGI
QSAEQKRTVAT"

complement(2648..2653)

/locus_tag="plu2375"

2826..3266

gene

CDS
 /locus_tag="plu2376"
 2826.._3266
 /locus_tag="plu2376"
 /function="Unknown"
 /note="unnamed protein product; Similar to probable membrane protein ydgk of Escherichia coli"
 /codon_start=1
 /transl_table=11
 /protein_id="CAE14690.1"
 /db_xref="GI:36785708"
 /translation="MYQNDYHRIINGVLAIPAAYLIMTFLSASVMALYIMTEPSONN
 MLAGNNHTLMALFSAITVAWVWCFTLWVLLKLFIRSKRPPIFIIMLVITVIALK
 TFAFSPIDOMALRALAMPILAAAFVPIKSRSHRVMTPTDR"
 3393.._3974
 /gene="rnfA"
 /locus_tag="plu2377"
 3393.._3974
 /gene="rnfA"
 /locus_tag="plu2377"
 /codon_start=1
 /transl_table=11
 /product="Electron transport complex protein RnfA"
 /protein_id="CAE14692.1"
 /db_xref="GI:36785709"
 /translation="MTDYFLFVGVTVNNVIVYKPLGLCPFGVSKLETATGMGPA
 TFEVMTLASVCSWLVNPFILPLDLYRTLSPIVAVYQTELVVRETSPILVRL
 LGIFLITTNCAVLGVALLINQSHDFLOSALYGGAAAGFSLVWVLFPAIRERLAV
 ADVAPAPKSSISGLITAGLSLAFMFGFSLVKP"
 3974.._4606
 /gene="rnfB"
 /locus_tag="plu2378"
 3974.._4606
 /gene="rnfB"
 /locus_tag="plu2378"
 /codon_start=1
 /transl_table=11
 /product="Electron transport complex protein RnfB"
 /protein_id="CAE14694.1"
 /db_xref="GI:36785710"
 /translation="WTSLSMTAIGALSALGILFGLIGFAARPOVEDPIVENIDIL
 POSOCGCGYPCGRPYAEVANNENMINKAPGEGVWMLTSELGIDPQLDGDNL
 ONPARKAFIDENICIGCTKIOACPDVAIVGATRAMHTVEDLCTGCDLCAVPCPTD
 CITMIPVATTTSMWKMDELKTIIPVANIIVPGSHASISIKIPIKIBAEHV"
 4687.._4692
 /gene="rnfC"
 /locus_tag="plu2379"
 4687.._4692
 /gene="rnfC"
 /locus_tag="plu2379"
 /codon_start=1
 /transl_table=11
 /product="Electron transport complex protein RnfC"
 /protein_id="CAE14696.1"
 /db_xref="GI:36785711"
 /translation="MPLFLNLINKNKIMDPAGGIHPPEMKLSSOSPRIATPPEELI
 IPIQCHLGTGELLVYKGVDEVLAKGQPIRTTGTATVPNASTSTIIAIEPCIRAHSG
 LKELCVLRPDGKQWGERIQTPYOSLDMDTLINRHQAGLGGAGPTAKIOS
 GRHGLKTLINAAECPEYITADRLMQEHANLELIGIRILEHLKBEQILIGIDNKP
 AAIKLTIVLKDGKSIIVRIVIPKYPSGAKOLKILITGKEVPSGSSSTIYIMONV
 GTVAIKRAIIDEPLIERVITIGEAETPGNFMALGTPOVLOLAGFAPOSEOM
 VINGPIMGFTLPDINVPITKISNCITLAPSHOMDSYTERACIRGLCYEACPARLL
 PQQLYWRSQGEHKKADHNLPCIEGCAATYCPSPITPVQYTRQKKAIRALDQEN
 RSTEARIRBAQARERERKLARDESHKSAVOVDKSAVDALSRVKEQANAG
 KVAALIEGQLPDNAAVIAARQARBEQIRAAQARQAVENNSITPESSTADEKRAL
 AAALARAKKAAQONSETSPPEARBEETITSEKODIKDAVAAALARVAKKACQOP
 LTFQAEETPPAEETDPRKAAVAAIARVAKKACQOTPLTETOETPPAEETDPRKAAV
 AAALARVAKKACQOTPLTETOETPPAEETDPRKAAVAAIARVAKKACQOTQLTETO

gene
 /gene="rnfD"
 /locus_tag="plu2380"
 7170.._8249
 /gene="rnfD"
 /locus_tag="plu2380"
 /codon_start=1
 /transl_table=11
 /product="Electron transport complex protein RnfD"
 /protein_id="CAE14698.1"
 /db_xref="GI:36785712"
 /translation="MKRPVQIDNKKLKYASSPETHNOOSTSRIMLVVLAIPGAV
 QYPSYGTLPQULLAVITALLASALSLRQATMPRLQDSSALLTGLIGISLPL
 APWMLVIGTAFAVITAKHLVGLGQNPFPNPAVGVVILISFPVOMTGMPLMILIA
 TOTPLDLSLMVIPTGQTPDGTLLQOQGFDSQATPLDSRTGLTHIDVLRQ
 ILOGSLAGIQMOWNTAVLVGGLILNRRVITWQIPVAFILVGLCALTSWLDPSRY
 SPPLQLFSGATWLGAFIATDPVSASTTPKGRVFGAMIGLLVWVIRVYGVGPDALIA
 PAVLANITVPLIDHYTOPRVYGHK"
 join(8249.._8254,8259.._8888)
 /gene="rnfG"
 /locus_tag="plu2381"
 8249.._8254
 /gene="rnfG"
 /locus_tag="plu2381"
 8259.._8888
 /gene="rnfG"
 /locus_tag="plu2381"
 /codon_start=1
 /transl_table=11
 /product="Electron transport complex protein RnfG"
 /protein_id="CAE14701.1"
 /db_xref="GI:36785713"
 /translation="MLETRRHGTTIATPAALTGTJAVVNSITKSTIAEQALQKS
 LLDQVPPALVNDIQNECYLVANALGNLPHRLYIARNGHPVAAALESTAPDYS
 GAIQILVAGDSFGKVLGVVTEHETRGDKIETRIISMNVFSGKITTSNDDQMA
 VKDGEFPOFTGATITTPRAVANSVRLTYIQTIPHLSSLESCGKS"
 8873.._8878
 /gene="rnfH"
 /locus_tag="plu2382"
 8873.._8878
 /gene="rnfH"
 /locus_tag="plu2382"
 8885.._9583
 /gene="rnfH"
 /locus_tag="plu2382"
 8885.._9583
 /gene="rnfH"
 /locus_tag="plu2382"
 /codon_start=1
 /transl_table=11
 /product="Electron transport complex protein RnfH"
 /protein_id="CAE14703.1"
 /db_xref="GI:36785714"
 /translation="MSETKMLFPIQGLKRNKNSALVQLGICPLLAVSSTATNLGLGLA

CDS
 4599.._7157
 /gene="rnfC"
 /locus_tag="plu2379"
 4599.._7157
 /gene="rnfC"
 /locus_tag="plu2379"
 /codon_start=1
 /transl_table=11
 /product="Electron transport complex protein RnfC"
 /protein_id="CAE14696.1"
 /db_xref="GI:36785711"
 /translation="MPLFLNLINKNKIMDPAGGIHPPEMKLSSOSPRIATPPEELI
 IPIQCHLGTGELLVYKGVDEVLAKGQPIRTTGTATVPNASTSTIIAIEPCIRAHSG
 LKELCVLRPDGKQWGERIQTPYOSLDMDTLINRHQAGLGGAGPTAKIOS
 GRHGLKTLINAAECPEYITADRLMQEHANLELIGIRILEHLKBEQILIGIDNKP
 AAIKLTIVLKDGKSIIVRIVIPKYPSGAKOLKILITGKEVPSGSSSTIYIMONV
 GTVAIKRAIIDEPLIERVITIGEAETPGNFMALGTPOVLOLAGFAPOSEOM
 VINGPIMGFTLPDINVPITKISNCITLAPSHOMDSYTERACIRGLCYEACPARLL
 PQQLYWRSQGEHKKADHNLPCIEGCAATYCPSPITPVQYTRQKKAIRALDQEN
 RSTEARIRBAQARERERKLARDESHKSAVOVDKSAVDALSRVKEQANAG
 KVAALIEGQLPDNAAVIAARQARBEQIRAAQARQAVENNSITPESSTADEKRAL
 AAALARAKKAAQONSETSPPEARBEETITSEKODIKDAVAAALARVAKKACQOP
 LTFQAEETPPAEETDPRKAAVAAIARVAKKACQOTPLTETOETPPAEETDPRKAAV
 AAALARVAKKACQOTPLTETOETPPAEETDPRKAAVAAIARVAKKACQOTQLTETO

Query Match 66.0%; Score 19.8; DB 1; Length 336609;
 Best Local Similarity 91.3%; Pred. No. 1.8e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTTAATTGACGACGATCAATTC 27
 Db 12713 CTTAAATTCACGATCAATTC 12735

RESULT 10
 LOCUS AX770906 349980 bp DNA linear PAT 02-JUL-2003
 DEFINITION Sequence 37 from Patent WO02094867.
 ACCESSION AX770906
 VERSION AX770906.1 GI:32438070
 KEYWORDS

SOURCE
ORGANISM Photorhabdus luminescens
Bacteri: Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.

REFERENCE
AUTHORS 1 Duchaud, E., Taourit, S., Glaeser, P., Frangeul, L., Kunst, F.,
Danchin, A. and Buchrieser, C.
TITLE Sequence of the Photorhabdus luminescens strain TT01 genome and
uses
JOURNAL Patent: WO 02094867-A 37 28-NOV-2002;
INSTITUT PASTEREUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
LOCATION/Qualifiers
FEATURES
source 1. 349980
/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/db_xref="taxon:29488"
/note="sequence length too big. Splitted into 2
sequences.-seq 00037: 1 to 349980-seq 10786: 300001 to
530312"

ORIGIN
Query Match 66.0%; Score 19.8; DB 6; Length 349980;
Best Local Similarity 91.3%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTTAATTGACGAGTACATTC 27
Db 110745 CTTAATTGACGAGTACATTC 110723

RESULT 11
AC100379 59008 bp DNA linear HTG 28-NOV-2002
LOCUS Mus musculus clone RP23-131C19, LOW-PASS SEQUENCE SAMPLING.
AC100379
AC100379.3 GI:25815482
VERSION HTG; HTGS PHAS80.
KEYWORDS Mus musculus (house mouse)
SOURCE
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 Birren, B., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-131C19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 59008)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearliano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,
Lamasares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhahang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 59008)

AUTHORS
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearliano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hages, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Mihova, T.,
Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mhova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunhahang, P., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N.,
Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL Submitted (28-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 28, 2002 this sequence version replaced gi:24962776.
ALL repeats were identified using RepeatMasker:
Smir, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L15213
Center clone name: 131_C_19

NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
725 824: contig of 724 bp in length
* 725 824: gap of 100 bp
* 825 1527: contig of 703 bp in length
* 1528 1627: gap of 100 bp
* 1628 2324: contig of 697 bp in length
* 2325 2424: gap of 100 bp
* 2425 3114: contig of 690 bp in length
* 3115 3214: gap of 100 bp
* 3215 3925: contig of 711 bp in length
* 3926 4025: gap of 100 bp
* 4026 4728: contig of 703 bp in length
* 4729 4828: gap of 100 bp
* 4829 5538: contig of 710 bp in length
* 5539 5638: gap of 100 bp
* 5639 6354: gap of 716 bp in length
* 6355 6454: gap of 100 bp
* 6455 7158: contig of 704 bp in length
* 7159 7258: gap of 100 bp
* 7259 7977: contig of 719 bp in length
* 7978 8077: gap of 100 bp
* 8078 8741: contig of 664 bp in length
* 8742 8841: gap of 100 bp
* 8842 9527: contig of 686 bp in length
* 9528 9627: gap of 100 bp
* 9628 10313: contig of 686 bp in length
* 10314 10413: gap of 100 bp
* 10414 11099: contig of 686 bp in length
* 11100 11199: gap of 100 bp

```

* 11200 11915: contig of 716 bp in length
* 11916 12015: gap of 100 bp in length
* 12016 12717: contig of 702 bp in length
* 12718 12817: gap of 100 bp in length
* 12818 13522: contig of 705 bp in length
* 13523 13622: gap of 100 bp in length
* 13623 14304: contig of 682 bp in length
* 14305 14404: gap of 100 bp in length
* 14405 15085: contig of 681 bp in length
* 15086 15185: gap of 100 bp in length
* 15186 15892: contig of 707 bp in length
* 15893 16685: contig of 693 bp in length
* 16686 16785: gap of 100 bp in length
* 16786 17467: contig of 682 bp in length
* 17468 17567: gap of 100 bp in length
* 17568 18269: contig of 702 bp in length
* 18270 18369: gap of 100 bp in length
* 18370 19092: contig of 723 bp in length
* 19093 19192: gap of 100 bp in length
* 19193 19917: contig of 725 bp in length
* 19918 20017: gap of 100 bp in length
* 20019 20719: contig of 702 bp in length
* 20720 20819: gap of 100 bp in length
* 20820 21521: contig of 702 bp in length
* 21522 21621: gap of 100 bp in length
* 21622 22295: contig of 674 bp in length
* 22296 22395: gap of 100 bp in length
* 22396 23087: contig of 692 bp in length
* 23088 23187: gap of 100 bp in length
* 23188 23889: contig of 702 bp in length
* 23890 23989: gap of 100 bp in length
* 23990 24708: contig of 719 bp in length
* 24709 24808: gap of 100 bp in length
* 24809 25427: contig of 619 bp in length
* 25428 25527: gap of 100 bp in length
* 25528 26230: contig of 703 bp in length
* 26231 26330: gap of 100 bp in length
* 26331 27004: contig of 674 bp in length
* 27005 27104: gap of 100 bp in length
* 27105 27800: contig of 696 bp in length
* 27801 27900: gap of 100 bp in length
* 27901 28595: contig of 695 bp in length
* 28596 28695: gap of 100 bp in length
* 28696 29354: contig of 659 bp in length
* 29355 29454: gap of 100 bp in length
* 29455 30162: contig of 708 bp in length
* 30163 30262: gap of 100 bp in length
* 30263 30953: contig of 691 bp in length
* 30954 31053: gap of 100 bp in length
* 31054 31766: contig of 713 bp in length
* 31767 31866: gap of 100 bp in length
* 31867 32567: contig of 701 bp in length
* 32568 33377: gap of 100 bp in length
* 33378 33477: gap of 100 bp in length
* 33478 34200: contig of 723 bp in length
* 34201 34300: gap of 100 bp in length
* 34301 35023: contig of 722 bp in length
* 35023 35122: gap of 100 bp in length
* 35123 35765: contig of 643 bp in length
* 35766 35865: gap of 100 bp in length
* 35866 36664: gap of 699 bp in length
* 36665 37371: contig of 707 bp in length
* 37372 37471: gap of 100 bp in length
* 37472 38173: contig of 702 bp in length
* 38174 38273: gap of 100 bp in length
* 38274 38967: contig of 694 bp in length
* 38968 39067: gap of 100 bp in length
* 39068 39772: contig of 705 bp in length
* 39773 39872: gap of 100 bp in length
* 39873 40562: contig of 690 bp in length

```

```

Query Match 65.3%; Score 19.6; DB 2; Length 5908;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 22; Conservativity 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACACTTATTGACCGACTTAATT 26
Db 9942 CACACTTATTGACCTAGTAATAATT 9967

RESULT 12
AC100388/c
LOCUS AC100388
DEFINITION Mus musculus clone RP23-133A3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC100388
VERSION AC100388.1 GI:17047754
KEYWORDS HTG; HTGS PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 60978)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-133A3
Unpublished
2 (bases 1 to 60978)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckgalter, B.,
Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heathford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lenoczky, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPeckers, R., Meldrum, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Recta, R., Rieback, M., Riley, R., Rie, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnpack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G.,
Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Genome code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

```
----- Project Information
Center Project name: L15250
Center Clone name: 133_A_3
-----
* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1      680: contig of 680 bp in length
*      681      780: gap of 100 bp in length
*      781      1490: contig of 710 bp in length
*      1491      1590: gap of 100 bp in length
*      1591      2312: contig of 722 bp in length
*      2313      2412: gap of 100 bp in length
*      2413      3099: contig of 687 bp in length
*      3100      3199: gap of 100 bp in length
*      3200      3899: contig of 700 bp in length
*      3900      3999: gap of 100 bp in length
*      4000      4692: contig of 693 bp in length
*      4693      4792: gap of 100 bp in length
*      4793      5495: contig of 703 bp in length
*      5496      5595: gap of 100 bp in length
*      5596      6310: contig of 715 bp in length
*      6311      6410: gap of 100 bp in length
*      6411      7105: contig of 695 bp in length
*      7106      7205: gap of 100 bp in length
*      7206      7933: contig of 728 bp in length
*      7934      8033: gap of 100 bp in length
*      8034      8719: contig of 686 bp in length
*      8720      8819: gap of 100 bp in length
*      8820      9539: contig of 720 bp in length
*      9540      9639: gap of 100 bp in length
*      9640      10367: contig of 728 bp in length
*      10368      10467: gap of 100 bp in length
*      10468      11158: contig of 691 bp in length
*      11159      11258: gap of 100 bp in length
*      11259      11965: contig of 707 bp in length
*      11966      12065: gap of 100 bp in length
*      12066      12758: contig of 693 bp in length
*      12759      12858: gap of 100 bp in length
*      12859      13557: contig of 699 bp in length
*      13558      13657: gap of 100 bp in length
*      13658      14362: contig of 705 bp in length
*      14363      14462: gap of 100 bp in length
*      14463      15183: contig of 721 bp in length
*      15184      15283: gap of 100 bp in length
*      15284      16051: contig of 768 bp in length
*      16052      16151: gap of 100 bp in length
*      16152      16861: contig of 710 bp in length
*      16862      16961: gap of 100 bp in length
*      16962      17661: contig of 700 bp in length
*      17662      17761: gap of 100 bp in length
*      17762      18448: contig of 687 bp in length
*      18449      18548: gap of 100 bp in length
*      18549      19259: contig of 711 bp in length
*      19260      19359: gap of 100 bp in length
*      19360      20087: contig of 728 bp in length
*      20088      20187: gap of 100 bp in length
*      20188      20876: contig of 689 bp in length
*      20877      20976: gap of 100 bp in length
*      20977      21670: contig of 694 bp in length
*      21671      21770: gap of 100 bp in length
*      21771      22420: contig of 650 bp in length
*      22421      22520: gap of 100 bp in length
*      22521      23227: contig of 707 bp in length
*      23227      23327: gap of 100 bp in length
*      23328

*      23328      24093: contig of 766 bp in length
*      24094      24193: gap of 100 bp in length
*      24194      24910: contig of 717 bp in length
*      24911      25010: gap of 100 bp in length
*      25011      25680: contig of 670 bp in length
*      25681      25780: gap of 100 bp in length
*      25781      26450: contig of 670 bp in length
*      26451      26550: gap of 100 bp in length
*      26551      27226: contig of 676 bp in length
*      27227      27326: gap of 100 bp in length
*      27327      28029: contig of 703 bp in length
*      28030      28129: gap of 100 bp in length
*      28130      28850: contig of 721 bp in length
*      28851      28950: gap of 100 bp in length
*      28951      29642: contig of 692 bp in length
*      29643      29742: gap of 100 bp in length
*      29743      30448: contig of 706 bp in length
*      30449      30548: gap of 100 bp in length
*      30549      31252: contig of 704 bp in length
*      31253      31352: gap of 100 bp in length
*      31353      32039: contig of 687 bp in length
*      32040      32139: gap of 100 bp in length
*      32140      32849: contig of 710 bp in length
*      32850      32949: gap of 100 bp in length
*      32950      33659: contig of 710 bp in length
*      33660      33759: gap of 100 bp in length
*      33760      34476: contig of 717 bp in length
*      34477      34576: gap of 100 bp in length
*      34577      35266: contig of 690 bp in length
*      35267      35366: gap of 100 bp in length
*      35367      36074: contig of 708 bp in length
*      36075      36174: gap of 100 bp in length
*      36175      36854: contig of 680 bp in length
*      36855      36954: gap of 100 bp in length
*      36955      37650: contig of 696 bp in length
*      37651      37750: gap of 100 bp in length
*      37751      38452: contig of 702 bp in length
*      38453      38552: gap of 100 bp in length
*      38553      39269: contig of 711 bp in length
*      39270      39369: gap of 100 bp in length
*      39370      40046: contig of 677 bp in length
*      40047      40146: gap of 100 bp in length
*      40147      40844: contig of 696 bp in length
*      40845      40944: gap of 100 bp in length
*      40945      41652: contig of 708 bp in length
*      41653      41752: gap of 100 bp in length
*      41753      42478: contig of 726 bp in length
*      42479      42578: gap of 100 bp in length
*      42579      43272: contig of 694 bp in length
*      43273      43372: gap of 100 bp in length
*      43373      44053: contig of 681 bp in length
*      44054      44153: gap of 100 bp in length
*      44154      44854: contig of 701 bp in length
*      44855      44954: gap of 100 bp in length
*      44955      45680: contig of 726 bp in length
*      45681      45780: gap of 100 bp in length
*      45781      46475: contig of 695 bp in length
*      46476      46575: gap of 100 bp in length
*      46576      47286: contig of 711 bp in length
*      47287      47386: gap of 100 bp in length
*      47387      48064: contig of 678 bp in length
*      48065      48164: gap of 100 bp in length
*      48165      48874: contig of 710 bp in length
*      48875      48974: gap of 100 bp in length
*      48975      49686: contig of 712 bp in length
*      49687      49786: gap of 100 bp in length
*      49787      50503: contig of 717 bp in length
*      50504      50603: gap of 100 bp in length
*      50604      51302: contig of 699 bp in length
*      51303      51402: gap of 100 bp in length
*      51403      52106: contig of 704 bp in length
*      52107      52206: gap of 100 bp in length
*      52207      52922: contig of 716 bp in length
```

* 52923 53022: gap of 100 bp
* 53023 53741: contig of 719 bp in length
* 53742 53841: gap of 100 bp
* 53842 54545: contig of 704 bp in length
* 54546 54645: gap of 100 bp
* 54646 55352: contig of 707 bp in length
* 55353 55452: gap of 100 bp

Query Match 65.3%; Score 19.6; DB 2; Length 60978;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 47184 ACTTCATTGACCTGATGATTTTCGC 47159

RESULT 13
AL358234
LOCUS
DEFINITION Human DNA sequence from clone RP11-164A7 on chromosome 10, complete
sequence.
ACCESSION AL358234
VERSION AL358234.14 GI:21211697
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 103930)
AUTHORS Phillimore, B.
JOURNAL Direct Submission
Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On May 25, 2002 this sequence version replaced gi:14269935.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-164A7 is from the library RP11-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.

FEATURES
source
1..103930
Location/Qualifiers

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-164A7"
/clone_lib="RP11-11.1"

ORIGIN
Query Match 65.3%; Score 19.6; DB 9; Length 103930;

Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 92999 CAGACTTAATTGACCGAGTACATT 93024

RESULT 14
AC026388
WPCOMMENT
Sequence split into 5 fragments

Fragment Name	Begin	End	LOCUS	AC026388	Accession	AC026388
AC026388.0	1	110000				
AC026388.1	100001	210000				
AC026388.2	200001	310000				
AC026388.3	300001	410000				
AC026388.4	400001	416492				

LOCUS AC026388 416492 bp DNA linear HTG 16-OCT-2001
DEFINITION Mus musculus chromosome 7 clone RP23-426B15, WORKING DRAFT
SEQUENCE, 15 unordered pieces.

ACCESSION AC026388
VERSION AC026388.17 GI:16118084
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 416492)
AUTHORS Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okunou, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooke, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Kocha, S.,
Ferreandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogulu, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasequez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Worley, K. and Gibbs, R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 416492)
AUTHORS Worley, K.C.
JOURNAL Direct Submission
Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 14, 2001 this sequence version replaced gi:12025600.

COMMENT
----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: MADM

Center clone name: RP23-426B15

----- Summary Statistics

Sequencing vector: M13, L08821

Chemistry: Dye-primer Bodipy: 75% of reads

Chemistry: Dye-terminator Big Dye: 25% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 413035 bases at least Q40

Consensus quality: 425796 bases at least Q20

Estimated insert size: 414396; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 6.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hosc.bcm.tmc.edu/hosc/genbank/draft_data.html)
 * NOTE: This sequence may represent more than one clone_
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 119764: contig of 119764 bp in length
 * 119765 119864: gap of unknown length
 * 119865 177594: contig of 57720 bp in length
 * 177595 177694: gap of unknown length
 * 177695 225261: contig of 47567 bp in length
 * 225262 225361: gap of unknown length
 * 225362 263833: contig of 38472 bp in length
 * 263834 263933: gap of unknown length
 * 263934 298288: contig of 34355 bp in length
 * 298289 298389: gap of unknown length
 * 298389 333082: contig of 34694 bp in length
 * 333083 333182: gap of unknown length
 * 333183 362065: contig of 28883 bp in length
 * 362066 362165: gap of unknown length
 * 362166 380755: contig of 18590 bp in length
 * 380756 380855: gap of unknown length
 * 380856 387924: contig of 7069 bp in length
 * 387925 388025: gap of unknown length
 * 388025 393894: contig of 5870 bp in length
 * 393895 393994: gap of unknown length
 * 393995 400042: contig of 6048 bp in length
 * 400043 400142: gap of unknown length
 * 400143 404818: contig of 4676 bp in length
 * 404819 404918: gap of unknown length
 * 404919 408719: contig of 3801 bp in length
 * 408720 408819: gap of unknown length
 * 408820 413398: contig of 4579 bp in length
 * 413399 413498: gap of unknown length
 * 413499 416492: contig of 2994 bp in length.

FEATURES
 source
 1. 416492
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="7"
 /clone="RP23-426B15"

ORIGIN
 Query Match 65.3%; Score 19.6; DB 2; Length 110000;
 Best Local Similarity 84.6%; Pred. No. 2.3e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGATGACATT 26
 |||||
 Db 24641 CACACTTCATTGACCTGATAATT 24666
 |||||

RESULT 15
 AC127467 139228 bp DNA linear MAM 03-MAY-2003
 LOCUS AC127467
 DEFINITION Atelerix albiventris clone LB4-4B15, complete sequence.
 AC127467
 ACCESSION AC127467
 VERSION AC127467.3 GI:30348997
 KEYWORDS
 SOURCE
 ORGANISM
 Atelerix albiventris (middle-African hedgehog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Insectivora; Erinaceidae; Erinaceinae;
 Atelerix.
 1 (bases 1 to 139228)
 Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Caraga,K.,

Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
 Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
 Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Latic,P., Lee-Lin,S.-Q.,
 Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
 Maskeri,B., McDowell,J., Peguirigan,C., Pearson,R., Portnoy,M.E.,
 Prasad,A., Reddi-Dugue,N., Schandler,K., Schuler,M.G., Shah,K.,
 Sison,C., Stancirpop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
 Vogt,J.L., Weheby,K.D., Wiggins,L., Young,A. and Green,E.D.
 NISC Comparative Sequencing Initiative
 JOURNAL Unpublished
 TITLE 2 (bases 1 to 139228)
 REFERENCES
 AUTHORS
 TITLE Direct Submission
 JOURNAL Submitted (17-JUL-2002) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 REFERENCE 3 (bases 1 to 139228)
 AUTHORS
 TITLE Direct Submission
 JOURNAL Submitted (22-AUG-2002) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 REFERENCE 4 (bases 1 to 139228)
 AUTHORS
 TITLE Direct Submission
 JOURNAL Submitted (03-MAY-2003) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 COMMENT
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_zoo@nhgri.nih.gov
 ----- Project Information
 Center project name: due
 Center clone name: 004B15

CLONE LENGTH: This sequence represents the entire insert of
 this clone unless otherwise noted. If there are overlapping
 clones, the overlaps are noted in the beginning and end of
 the Features section.
 Location/Qualifiers
 source 1. 139228
 /organism="Atelerix albiventris"
 /mol_type="genomic DNA"
 /db_xref="taxon:9368"
 /clone="LB4-4B15"
 /clone_1ib="LB4"
 /clone_1ibb="LB4"
 34774..34860
 misc_feature
 /note="single clone coverage"
 139221..139228
 /note="single clone coverage"

ORIGIN
 Query Match 65.3%; Score 19.6; DB 4; Length 139228;
 Best Local Similarity 84.6%; Pred. No. 2.3e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGATGACATT 26
 |||||
 Db 107135 CACACTTAATTGCCCGATGACATT 107160
 |||||

Search completed: July 20, 2004, 00:54:40
 Job time : 751.574 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: July 20, 2004, 00:15:28 ; Search time 164.754 Seconds
(without alignments)
773.552 Million cell updates/sec

Title: US-09-831-272-2

Perfect score: 30

Sequence: 1 cacactaatcgcgcgagtaacatccgc 30

Scoring table: IDENTITY_NUC

Gapex 10.0 , Gapex 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_290a04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	30	AAA27965	Aaa27965 Box W1 we
2	19.8	66.0	110000	7 ACFe7367_07	Continuation (8 of
3	19.8	66.0	110000	7 ACFe5384_1	Continuation (2 of
4	19.6	65.3	556	ACD94807	Human col
5	19.4	64.7	61	AAO98799	External
6	19.2	64.0	18753	4 ABL20682	Drosophila
7	19	63.3	14968	3 AAA55221	Human ade
8	19	63.3	14968	3 AAF21343	Human low
9	19	63.3	14968	4 AAD14464	Human IL-
10	19	63.3	14968	4 AAD15838	Human int
11	19	63.3	14968	7 ABZ97037	Human nuc
12	19	63.3	17844	3 AAA55223	Human ade
13	19	63.3	17904	3 AAF21345	Human low
14	19	63.3	17904	7 ABZ97039	Human nuc
15	19	63.3	33053	6 ABQ67005	Human ang
16	19	63.3	110000	6 ABQ67196	Listeria
17	19	63.3	110000	6 ABO69245_24	Continuation (25 o
18	18.8	62.7	1782	3 AAA96679	Nucleoid
19	18.8	62.7	1988	7 ADA69244	Arabidops
20	18.8	62.7	2348	3 AAA96678	Nucleoid
21	18.6	62.0	868	3 AAA52605	Eosinophi
22	18.6	62.0	945	4 AAL26424	Human bre
23	18.4	61.3	1506	5 AAS70860	DNA encod

c 24	18.4	61.3	3103	5 AAS85091	AAS85091 DNA encod
c 25	18.4	61.3	7500	6 ABL33115	Human imm
c 26	18.4	61.3	8361	4 ABL05662	Drosophila
c 27	18.4	61.3	8869	4 ABL13210	Drosophila
c 28	18.4	61.3	65608	6 ABL62910	Breast ca
c 29	18.4	61.3	65608	6 ABL64414	Stomach c
c 30	18.4	61.3	65608	6 ABL67668	Oesophagu
c 31	18.2	60.7	11658	4 ABL17680	Drosophila
c 32	18	60.0	403	6 ABL23600	Human ORF
c 33	18	60.0	468	3 AAC02933	Human sec
c 34	18	60.0	782	4 AAH08439	Human CDN
c 35	18	60.0	4413	4 ABL04492	Drosophila
c 36	18	60.0	5544	6 ABL34621	Human met
c 37	18	60.0	5544	6 ABL70478	Chemical
c 38	18	60.0	5544	6 AAS61441	Human gen
c 39	18	60.0	5787	7 AAL52041	Broccoli
c 40	18	60.0	6820	4 ABL12056	Drosophila
c 41	18	60.0	6901	4 ABL12076	Drosophila
c 42	18	60.0	23761	6 ABO80552	Human Can
c 43	17.8	59.3	534	4 AAL12387	Human bre
c 44	17.8	59.3	797	4 AAL21263	Human bre
c 45	17.8	59.3	10594	3 AAZ49991	Plasmid p

ALIGNMENTS

RESULT 1	AAA27965	standard; DNA; 30 BP.
ID	AAA27965	
XX	AAA27965;	
AC		
XX		
DT	15-AUG-2000	(first entry)
XX		
DE	Box W1 weak elicitor-responsive cis-element nucleotide sequence.	
XX		
KW	Box W1; elicitor-responsive cis-element; parsley; P1 promoter; ss;	
KW	chimeric promoter; pathogen infection; transgenic plant; resistance;	
KW	herbicide; local response; genetic engineering; disease resistant crop.	
XX		
OS	Petroleumum crispum.	
XX		
PN	MO200029592-A2.	
XX		
PD	25-MAY-2000.	
XX		
PF	12-NOV-1999; 99WC-EP008710.	
XX		
PR	12-NOV-1998; 98EP-00121160.	
XX		
PR	27-AUG-1999; 99EP-00116981.	
XX		
PA	(PLAC) MAX PLANCK GRS FOERDERUNG.	
XX		
PI	Kirsch C, Logemann E, Hahlbrock K, Ruehlon P, Somasich I;	
XX		
XX	WPI, 2000-387804/33.	
XX		
PT	Chimeric promoters mediating gene expression in plants upon pathogen	
PT	infection, useful for transgenic plant production comprises at least one	
PT	cis-acting element to direct elicitor-specific expression.	
XX		
XX	Claim 2; Page 32; 73pp; English.	
XX		
XX	This sequence represents Box W1, a weak elicitor-responsive cis-element	
CC	from the Parsley P1 promoter. The present invention relates to chimeric	
CC	promoters capable of mediating local gene expression in plants upon	
CC	pathogen infection. The chimeric promoters comprise at least one cis-	
CC	element (see AAA27964-A27979) capable of directing elicitor-specific	
CC	expression, and a minimal promoter. The chimeric promoters are useful for	
CC	producing a transgenic plant which has attained resistance or improved	
CC	resistance against a pathogen. The cis-acting element, chimeric promoter,	
CC	recombinant gene encoding the chimeric promoter, vector comprising the	

CC chimeric promoter and a compound capable of activating the chimeric
 CC promoter are useful for producing compound resistant plants, and for
 CC identifying and/or producing compounds capable of conferring induced
 CC resistance to a pathogen in a plant. A compound which specifically
 CC activates or inhibits genes activated in a plant when attacked by a
 CC pathogen is also useful as a plant protective agent or a herbicide. The
 CC chimeric promoter provides rapid and local response to pathogen attack
 CC but shows negligible activity in uninfected parts of the plants and
 CC therefore is most suitable for the engineering of disease resistant crops
 .XX

Sequence 30 BP; 9 A; 9 C; 4 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACACTTAATTGACCGAGTAACATTGCGC 30
 DB 1 CACACTTAATTGACCGAGTAACATTGCGC 30

RESULT 2

ACF67367_07/c
 Continuation (8 of 57) of ACF67367 from base 700001 (Photorhabdus luminescens nucleotide
 WP Sequence split into 57 fragments LOCUS ACF67367 Accession Acf67367

WP	Fragment Name	Begin	End
WP	ACF67367_01	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000

WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894

Query Match 66.0%; Score 19.8; DB 7; Length 110000;
 Best Local Similarity 91.3%; Pred. No. 52;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTTAATTGACCGAGTAACATTG 27
 DB 10530 CTTAATTGACCGAGTAACATTG 10508

RESULT 3

ACF65384_1/c
 Continuation (2 of 6) of ACF65384 from base 100001 (Photorhabdus luminescens nucleotide
 WP Sequence split into 6 fragments LOCUS ACF65384 Accession Acf65384

WP	Fragment Name	Begin	End
WP	ACF65384_1	1	110000
WP	ACF65384_2	100001	210000
WP	ACF65384_3	200001	310000
WP	ACF65384_4	300001	410000
WP	ACF65384_5	400001	510000
WP	ACF65384_5	500001	530312

Query Match 91.3%; Score 19.8; DB 7; Length 110000;
 Best Local Similarity 91.3%; Pred. No. 52;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CTTAATTGACCGAGTAACATTG 27
 DB 10745 CTTAATTGACCGAGTAACATTG 10723

RESULT 4

ACD94807/c
 ID ACD94807 standard; cDNA; 556 BP.

AC ACD94807;

XX 23-SEP-2003 (first entry)

DB Human colon cancer cell expressed cDNA #3219.

KW Open reading frame detection; genome sequencing; colon cancer;
 KW breast cancer; population genome analysis; genetic shift; cancer;
 KW antibiotic resistance; antibiotic non-tolerance; congenital disease;
 KW agriculture; food crop genome; resistance gene; retrovirus;
 KW influenza virus; eukaryotic pathogen detection; trypanosome; Plasmodium;
 KW gene; ss.

XX Homo sapiens.

XX US2002155438-A1.

XX 24-OCT-2002.

XX 27-SEP-1999; 99US-00406117.

XX 20-NOV-1998; 98US-00196716.

XX (SIMP/) SIMPSON A J G.

XX (NETO/) NETO E D.

XX (BREN/) BRENTANI R R.

PI Simpson AJG, Neto ED, Brentani RR;


```

XX WPI; 2003-182626/18.
DR
XX Determining open reading frames of genome of an organism e.g. a human
PT suffering from cancer involves use of single oligonucleotide primer at
PT low stringency for preparing single-stranded cDNA from mRNA of
PT individual.
XX
PS Example 9; Page 473; 959pp; English.
XX
CC The invention describes a method of determining open reading frames in
CC the genome of organism, comprising contacting mRNA from cell of organism
CC with a single oligonucleotide primer (I) at low stringency, preparing
CC single-stranded cDNA by reverse transcribing mRNA with (II), amplifying
CC cDNA, sequencing the product, and repeating the contacting, preparing
CC and amplifying steps with different primers and sequencing resulting
CC nucleic acids. The method is useful for: determining that a known
CC nucleotide sequence from a genome of an organism corresponds to a
CC nucleotide sequence of an open reading frame; for preparing a contig;
CC nucleic acid molecule from a genome of an organism; and for sequencing
CC all or part of a genome of an organism. mRNA is obtained from mammalian
CC or human cell which is associated with a pathological condition e.g. a
CC colon cancer or breast cancer cell. The method is useful for analyses of
CC populations of subjects and can be used to carry out genetic analyses of
CC large or small populations. further, it can be used to study living
CC systems to determine if, e.g. there have been genetic shifts which render
CC an individual or population more or less likely to be afflicted with
CC diseases such as cancer, to determine antibiotic resistance or non-
CC tolerance, and so forth. The method can also be used in the study of
CC congenital diseases, and the risk of affliction to a foetus, as well as
CC the study of whether the conditions are likely to be passed to offspring
CC through ova or sperm. The analyses for pathological conditions can be
CC carried out in all animals, plants, birds, fish, etc. Using this method,
CC in the area of agriculture, for example the genomes of food crops can be
CC studied to determine if resistance genes are present, defects in plant
CC genomes can also be studied in this way. Similarly, the method permits
CC determination of the pathogens which integrate into the genome, such as
CC retroviruses and other integrating viruses such as influenza virus, have
CC undergone shifts or mutations, which may require different approaches to
CC therapy. This method is also applied to eukaryotic pathogens, such as
CC trypanosomes, different types of Plasmodium, etc. The method essentially
CC eliminates sequencing of non-coding portions. This sequence represents a
CC polynucleotide isolated from human colon cancer cell cDNA library
XX
SQ Sequence 556 BP; 206 A; 75 C; 101 G; 174 T; 0 U; 0 Other;
XX
Query Match 65.3%; Score 19.6; DB 7; Length 556;
Best Local Similarity 84.6%; Pred. No. 29;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
QY 1 CACACTTAATTGACCGAGTAACATT 26
DB 263 CAGACTTAATTGTCGCCGAAATTAACATT 238
XX
RESULT 5
AAQ98799
ID AAQ98799 standard; cRNA; 61 BP.
XX
AC AAQ98799;
XX
DT 29-AUG-1996 (first entry)
XX
DE External guide sequence EGS-17 for targeting RNase P to CAT mRNA.
XX
KM external guide sequence; EGS; messenger RNA cleavage;
KM chloramphenicol acetyltransferase; RNase P; ribonuclease; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT stem_loop 1..11
FT /*tag= a

```

```

FT FT /label= D_stem-loop
FT stem_loop 12..28
FT /*tag= b
FT /label= anticodon_stem-loop
FT misc_structure 29..33
FT /*tag= c
FT /label= V_loop
FT stem_loop 34..50
FT /*tag= d
FT /label= T_stem-loop
FT misc_structure 51..61
FT /*tag= e
FT /label= amino_acyl_acceptor_stem
XX
PN MO9524489-A1.
XX
PD 14-SEP-1995.
XX
PF 07-MAR-1995; 95MO-US002816.
XX
PR 07-MAR-1994; 94US-00207547.
XX
PR 18-MAR-1994; 94US-00215082.
XX
PA (UYVA ) UNITV YALE.
XX
PI Yuan Y, Guerrier-Takada C, Altman S, Liu F;
XX
DR WPI; 1995-328280/42.
XX
PT Targeted ribonuclease P cleavage of RNA using an oligo:nucleotide -
PT comprising a target recognition sequence and a RNase P binding sequence,
PT useful for treating cancers and viral and bacterial infections.
XX
PS Example 5; Page 34; 94pp; English.
XX
CC Any RNA can be targeted for cleavage by RNase P, using a suitably
CC designed oligonucleotide as "external guide sequence" (EGS) to form a
CC hybrid with the target RNA and create a substrate for RNase P cleavage.
CC The EGSs contain sequences which are complementary to the target RNA and
CC which form secondary and tertiary structure similar to portions of a RNA
CC molecule. An EGS must contain at least 7 nucleotides which base pair with
CC the target sequence 3' to the intended cleavage site to form a structure
CC like the amino acyl acceptor stem, nucleotides which base pair to form a
CC stem and loop similar to the T stem-loop, followed by at least 3
CC nucleotides that base pair with the target sequence to form a structure
CC like the dihydroxyuracil stem. In an example, chimeric mRNA-EGS
CC substrates for guiding RNase P to chloramphenicol acetyltransferase (CAT)
CC mRNA were identified by an in vitro selection procedure. The present
CC sequence is the EGS segment of such a chimeric mRNA-EGS (CAT) substrate
XX
SQ Sequence 61 BP; 17 A; 20 C; 13 G; 0 T; 11 U; 0 Other;
XX
Query Match 64.7%; Score 19.4; DB 2; Length 61;
Best Local Similarity 58.6%; Pred. No. 26;
Matches 17; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
XX
QY 2 ACACTTAATTGACCGAGTAACATTGCC 30
DB 16 ACACTTAATTGACCGAGTAACATTGCC 44
XX
RESULT 6
ABL20682/c
ID ABL20682 standard; DNA; 18753 BP.
XX
AC ABL20682;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 13519.
XX
KM Drosophila developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ds.

```

```

XX OS Drosophila melanogaster.
XX PN MO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX DR WPI, 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Claim 1, SEQ ID NO 13519; 21bp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signaling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB16176-ABJ30511), expressed DNA
XX CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
XX CC AB872072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 18753 BP; 5401 A; 4009 C; 3969 G; 5374 T; 0 U; 0 Other;
QY Query Match 64.0%; Score 19.2; DB 4; Length 18753;
Db Best Local Similarity 87.5%; Pred. No. 76;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 CACTTAATTGACCGAGTAACATT 26
Db 2327 CACTTAATTGACCGAGTAACATT 2304
|||||
AAAS5221/c
ID AAAS5221 standard; DNA; 14968 BP.
XX AC AAAS5221;
XX AC AAAS5221;
XX DT 28-JUL-2000 (first entry)
XX DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:95.
XX OS Human adenosine receptor; low adenosine antisense oligonucleotide;
XX OS phosphorothioate; impaired respiration; inflammation; allergy;
XX OS allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
XX OS antiasthmatic; cytostatic; analgesic; impaired airway;
XX OS lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
XX OS respiratory distress syndrome; pain; cystic fibrosis; emphysema;
XX OS pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
XX OS cancer; leukemia; lymphoma; carcinoma; metastasis; BS.
XX OS Homo sapiens.
XX OS MO200009525-A2.
XX PN 24-FEB-2000.
XX PD 03-AUG-1999; 99WO-US017712.
XX PF
XX

```

```

PR 03-AUG-1998; 98US-0095212P.
XX PA (UYEC-) UNIV EAST CAROLINA.
XX PI Nyce JW;
XX DR WPI, 2000-205971/18.
XX PT New antisense oligonucleotides useful for treating e.g. pulmonary
XX PT vasoconstriction, inflammation, allergies, asthma, hypertension,
XX PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
XX PT cancers.
XX PS Disclosure; Page 1260-1264; 1343pp; English.
XX CC The present invention describes a new composition comprising an antisense
XX CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
XX CC nucleic acids involved in bronchoconstriction, allergies, and/or
XX CC inflammation. The ON can have antiinflammatory, antiasthmatic,
XX CC antiasthmatic, cytostatic and analgesic activities. The compositions are
XX CC useful for the treatment of diseases associated with inflammation,
XX CC impaired airways, including lung disease and diseases whose secondary
XX CC effects afflict the lungs of a subject. They can be used for treating
XX CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
XX CC impaired respiration, respiratory distress syndrome, pain, cystic
XX CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
XX CC carcinomas, and cancers which may metastasize to the lungs, including
XX CC breast and prostate cancer. The reduction of the adenosine content of the
XX CC ONs reduces side effects. The A-containing ONs break down with the
XX CC release of deoxyadenosine which activates adenosine receptors causing the
XX CC bronchoconstriction and inflammation. AA32313 to AA35312 represent the
XX CC nucleotide sequences given in the sequence listing from the present
XX CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
XX CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
XX CC from the previously named sequences. SEQ ID NO:11 to 1680 (AA32323 to
XX CC AA33992) are specifically claimed ONs from the present invention. N.B.
XX CC Sequences given in the disclosure of the present invention do not match
XX CC up with their corresponding SEQ ID NO: sequences given in the sequence
XX CC listing
SQ Sequence 14968 BP; 5109 A; 2436 C; 2757 G; 4666 T; 0 U; 0 Other;
QY Query Match 63.3%; Score 19; DB 3; Length 14968;
Db Best Local Similarity 81.5%; Pred. No. 91;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 AACTTAATTGACCGAGTAACATTG 28
Db 2190 AACTTAATTGACCGAGTAACATTG 2164
|||||
AAAF21343/c
ID AA21343 standard; DNA; 14968 BP.
XX AC AA21343;
XX AC AA21343;
XX DT 14-MAR-2001 (first entry)
XX DE Human low adenosine antisense oligonucleotide related sequence #2910.
XX OS Human low adenosine antisense oligonucleotide; phosphorothioate; allergy;
XX OS human; airway disorder; bronchoconstriction; lung inflammation;
XX OS surfactant depletion; respiratory; bronchodilator; antiinflammatory;
XX OS immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
XX OS respiratory obstruction; pulmonary obstruction; impaired respiration;
XX OS surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
XX OS respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
XX OS pulmonary hypertension; emphysema; pulmonary transplantation rejection;
XX OS chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
XX OS cancer; BS.
XX

```

OS Homo sapiens.
 XX W0200062736-A2.
 PN 26-OCT-2000.
 PD 24-MAR-2000; 2000MO-US008020.
 PF 06-APR-1999; 99US-0127958P.
 PR (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 PI Nyce JW;
 XX WPI; 2000-679539/66.
 DR Low adenosine (A) content antisense oligonucleotides which do not trigger
 PT adenosine receptors during metabolism, useful e.g. for treating cancers
 PT and respiratory obstructions.
 XX
 PS Disclosure; Page 1345-1348; 1592pp; English.
 XX
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiaesthetic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulin and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokines,
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
 CC surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention
 CC
 SQ Sequence 14968 BP; 5108 A; 2436 C; 2758 G; 4666 T; 0 U; 0 Other;
 Query Match 63.3%; Score 19; DB 3; Length 14968;
 Best Local Similarity 81.5%; Pred. No. 91;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

XX Human, interleukin 15; IL-15; gene therapy; chromosome 4q31; infection;
 KW drug screening; anthropological lineage; paternity testing; HIV;
 KW Human Immunodeficiency Virus; forensic application; T-cell leukaemia; ds.
 XX Homo sapiens.
 XX
 FH Key
 FT intron
 FT
 FT Location/Qualifiers
 FT 1..346
 FT /*tag= b
 FT /cons_splice= (5'site:NO, 3'site:NO)
 FT /number= 1
 FT replace(101, T)
 FT /*tag= o
 FT replace(191, C)
 FT /*tag= p
 FT replace(192, C)
 FT /*tag= q
 FT replace(267, T)
 FT /*tag= x
 FT replace(333, T)
 FT /*tag= b
 FT 347..13604
 FT /product= "Human IL-15 protein"
 FT 347..357
 FT /*tag= a
 FT /number= 1
 FT /*tag= c
 FT 358..1350
 FT /*tag= d
 FT /cons_splice= (5'site:NO, 3'site:NO)
 FT /number= 2
 FT replace(367, A)
 FT /*tag= t
 FT replace(431, T)
 FT /*tag= u
 FT replace(504, C)
 FT /*tag= v
 FT 1351..1448
 FT /*tag= e
 FT /number= 2
 FT 1449..2806
 FT /*tag= f
 FT /cons_splice= (5'site:NO, 3'site:NO)
 FT /number= 3
 FT replace(1536, A)
 FT /*tag= w
 FT 2807..2892
 FT /*tag= g
 FT /number= 3
 FT 2893..8695
 FT /*tag= h
 FT /number= 4
 FT 8696..8740
 FT /*tag= i
 FT /number= 4
 FT 8741..10602
 FT /*tag= j
 FT /number= 5
 FT replace(8866, A)
 FT /*tag= x
 FT replace(10504, G)
 FT /*tag= y
 FT 10603..10739
 FT /*tag= k
 FT /number= 5
 FT 10740..13492
 FT /*tag= l
 FT /number= 6
 FT /cons_splice= (5'site:NO, 3'site:NO)
 FT 13493..13604
 FT /*tag= m
 FT /number= 6

FT intron 13605..14968
 FT /*tag= n
 FT /number= 7
 FT /cons_splice= (5'site:NO, 3'site:NO)
 PN MO200158914-A2.
 OS
 PD 16-AUG-2001.
 PF 08-FEB-2001; 2001MO-US004130.
 PR 08-FEB-2000; 2000US-0181059P.
 PA (GENA-) GENAISSANCE PHARM INC.
 PI Anastasio AE, Chew A, Denton RR, Nandabalan K, Stephens JC;
 DR WPI, 2001-522460/57.
 DR P-PSDB; AAE08576.
 XX
 XX
 PT Novel polynucleotides comprising one of 11, P31-PS11, single nucleotide
 PT polymorphisms in human interleukin-15 gene, and useful for treating
 PT disorders affected by expression of function of interleukin-15 isogene.
 PS
 PS Claim 19; Fig 1; 78pp; English.
 XX
 XX The present sequence is human interleukin-15 (IL-15) gene allele located
 CC on chromosome 4q31. The polymorphic variants of IL-15 genes are useful
 CC for studying the expression and function of IL-15 and expressing IL-15
 CC protein for use in useful for screening for candidate drugs to treat
 CC diseases related to IL-15 activity. Genotyping or haplotyping an
 CC individual at the novel IL-15 polymorphic sites are useful for studying
 CC population diversity, anthropological lineage, the significance of
 CC diversity and lineage of the phenotypic level, paternity testing,
 CC forensic applications and for identifying associations between IL-15
 CC genetic variation and a trait such as level of drug response or
 CC susceptibility to disease. Identifying an association between a genotype
 CC or haplotype and a trait, is useful for developing diagnostic tests and
 CC therapeutic treatments for infections, human immunodeficiency virus and T
 CC -cell leukemia. The identification of an association between a clinical
 CC response and a genotype or haplotype (or haplotype pair) for the IL-15
 CC gene may be the basis for designing a diagnostic method to determine
 CC those individuals who will or will not respond to the treatment, or
 CC alternatively, will respond at a lower level and thus may require more
 CC treatment, i.e. a greater dose of a drug. The genotyping or haplotyping
 CC methods are also useful for developing drugs targeting IL-15. The
 CC genotyping and haplotyping methods are also useful in designing clinical
 CC trials. IL-15 DNA is useful for therapeutic purposes for treating
 CC disorders affected by expression of function of novel IL-15 isogene and
 CC also in gene therapy. Expression of an IL-15 isogene may be turned off by
 CC transforming a targeted organ, tissue or cell population of an
 CC expression vector that expresses high levels of untranslatable mRNA for
 CC the isogene
 XX
 SQ Sequence 14968 BP; 5108 A; 2436 C; 2758 G; 4666 T; 0 U; 0 Other;
 Query Match 63.3%; Score 19; DB 4; Length 14968;
 Best Local Similarity 81.5%; Pred. No. 91;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 ACACCTTAATTGACCGAGTAACATTCG 28
 DB 2190 ACACCTAATGACCAAGTAACATTCG 2164
 RESULT 10
 AAD15838/c
 ID AAD15838 standard; DNA; 14968 BP.
 XX
 AC AAD15838;
 XX
 DT 15-NOV-2001 (first entry)
 ,XX

DE Human interleukin 15 (IL-15) gene.
 XX
 XX Human; interleukin 15; IL-15; gene therapy; chromosome 4q31; infection;
 KM drug screening; anthropological lineage; paternity testing; HIV;
 KM Human Immunodeficiency Virus; forensic application; T-cell leukemia; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 101
 FT /*tag= a
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 191
 FT /*tag= b
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 192
 FT /*tag= c
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 267
 FT /*tag= d
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 333
 FT /*tag= e
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 367
 FT /*tag= f
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 431
 FT /*tag= g
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 504
 FT /*tag= h
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 1536
 FT /*tag= i
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 8866
 FT /*tag= j
 FT /note= "This degenerate base represents polymorphic site"
 FT misc_feature 10504
 FT /*tag= k
 FT /note= "This degenerate base represents polymorphic site"
 XX
 PN MO200158914-A2.
 PD 16-AUG-2001.
 PF 08-FEB-2001; 2001MO-US004130.
 PR 08-FEB-2000; 2000US-0181059P.
 PA (GENA-) GENAISSANCE PHARM INC.
 PI Anastasio AE, Chew A, Denton RR, Nandabalan K, Stephens JC;
 DR WPI, 2001-522460/57.
 XX
 XX Novel polynucleotides comprising one of 11, P31-PS11, single nucleotide
 PT polymorphisms in human interleukin-15 gene, and useful for treating
 PT disorders affected by expression of function of interleukin-15 isogene.
 PS
 PS Disclosure; Page 73-78; 78pp; English.
 XX
 XX The present sequence is human interleukin-15 (IL-15) gene located on
 CC chromosome 4q31. The polymorphic variants of IL-15 genes are useful for
 CC studying the expression and function of IL-15 and expressing IL-15
 CC protein for use in useful for screening for candidate drugs to treat
 CC diseases related to IL-15 activity. Genotyping or haplotyping an
 CC individual at the novel IL-15 polymorphic sites are useful for studying
 CC population diversity, anthropological lineage, the significance of
 CC diversity and lineage of the phenotypic level, paternity testing,
 CC forensic applications and for identifying associations between IL-15
 CC genetic variation and a trait such as level of drug response or

CC susceptibility to disease. Identifying an association between a genotype
CC or haplotype and a trait is useful for developing diagnostic tests and
CC therapeutic treatments for infections, human immunodeficiency virus and T
CC -cell leukemia. The identification of an association between a clinical
CC response and a genotype or haplotype (or haplotype pair) for the IL-15
CC gene may be the basis for designing a diagnostic method to determine
CC those individuals who will or will not respond to the treatment, or
CC alternatively, will respond at a lower level and thus may require more
CC treatment, i.e. a greater dose of a drug. The genotyping or haplotyping
CC methods are also useful for developing drugs targeting IL-15. The
CC genotyping and haplotyping methods are also useful in designing clinical
CC trials. IL-15 DNA is useful for therapeutic purposes for treating
CC disorders affected by expression of function of novel IL-15 isogene and
CC also in gene therapy. Expression of an IL-15 isogene may be turned off by
CC transforming a targeted organ, tissue or cell population of an
CC expression vector that expresses high levels of untranslatable mRNA for
CC the isogene.

SQ Sequence 14968 BP; 5107 A; 2432 C; 2755 G; 4663 T; 0 U; 11 Other;

Query Match 63.3%; Score 19; DB 4; Length 14968;
Best Local Similarity 81.5%; Pred. No. 91;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 ACACCTAATTGACCGAGTACATTGG 28
DB 2190 ACACTTACTGACCAAGTACATTGG 2164

RESULT 11
AB297037/c
ID AB297037 standard; DNA; 14968 BP.

AC AB297037;
DT 17-OCT-2003 (first entry)

DE Human nucleic acid sequence.

KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiaesthetic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.

OS Homo sapiens.

PN MO200285308-A2.

PD 31-OCT-2002.

PF 23-APR-2002; 2002WO-US01135.

PR 24-APR-2001; 2001US-0286137P.

PA (EPIC-) EPITGENESIS PHARM INC.

PI Myce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;

DR WPI; 2003-229219/22.

PT Pharmaceutical composition for treating ailments associated with impaired
PT respiration, has oligo(s) antisense to specific gene(s) or its
PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT ubiquinone.

PS Disclosure; SEQ ID NO 12279; 872pp; English.

CC The invention relates to a novel pharmaceutical composition, which has a
CC first active agent comprising an oligonucleotide antisense to the
CC initiation codon, coding region, 5' or 3' end genomic flanking regions,

CC 5' and 3' intion-exon junctions, or regions within 2-10 nucleotides of
CC junctions of genes encoding a polypeptide associated with lung and/or
CC nasal airway dysfunction and a second active agent comprising an
CC antiinflammatory steroid and ubiquinone. A composition of the invention
CC has antiinflammatory, antiallergic, antiaesthetic, hypotensive,
CC immunosuppressive, and cytostatic activity. The composition may have a
CC use in antisense gene therapy. The composition is useful for treating or
CC preventing a respiratory, lung or malignant disease or condition, also
CC for enhancing the prophylactic or therapeutic respiratory effect of an
CC antiinflammatory steroid in a subject, for reducing or depleting levels
CC of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC receptor, producing bronchodilation, increasing levels of ubiquinone or
CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC lung inflammation, lung allergies, or a respiratory disease or condition.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 14968 BP; 5108 A; 2436 C; 2758 G; 4666 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 7; Length 14968;
Best Local Similarity 81.5%; Pred. No. 91;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 ACACCTAATTGACCGAGTACATTGG 28
DB 2190 ACACTTACTGACCAAGTACATTGG 2164

RESULT 12
AAA5223/c
ID AAA5223 standard; DNA; 17844 BP.

AC AAA5223;
DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:97.

KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiaesthetic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

PN MO200009525-A2.

PD 24-FEB-2000.

PF 03-AUG-1999; 99WO-US017712.

PR 03-AUG-1998; 98US-0095212P.

PA (UYEC-) UNITV EAST CAROLINA.

PI Myce JW;

DR WPI; 2000-205971/18.

PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers.

PS Disclosure; Page 1264-1268; 1343pp; English.

CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets

CC nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammatory. The ON can have antiinflammatory, antiallergic,
 CC antistatic, cyostatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
 CC impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of the
 CC ON reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing the
 CC bronchoconstriction and inflammation. AAA3233 to AAA3532 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
 CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
 CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA3233 to
 CC AAA33992) are specifically claimed ONs from the present invention. N.B.
 CC Sequences given in the disclosure of the present invention do not match
 CC up with their corresponding SEQ ID NO: sequences given in the sequence
 CC listing

SQ Sequence 17844 BP; 6000 A; 2932 C; 3322 G; 5590 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 3; Length 17844;

Best Local Similarity 81.5%; Pred. No. 93; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACACCTTAATTGACCGAGTACATTG 28

4580 ACACCTTAATTGACCGAGTACATTG 4554

RESULT 13
 AAF21345/c

ID AAF21345 standard; DNA; 17904 BP.

XX AAF21345;

DT 14-MAR-2001 (first entry)

XX Human low adenosine antisense oligonucleotide related sequence #2912.

XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW immunosuppressive; antistatic; analgesic; hypotensive; cyostatic;
 KW surfactant hypoproduction; pulmonary obstruction; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

XX Homo sapiens.

XX WO200062736-A2.

XX 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US008020.

XX 06-APR-1999; 99US-0127958P.

XX (UYEC-) UNIV EAST CAROLINA.

XX (NYCE/) NYCE J W.

XX NYCE JW;

XX WPI, 2000-679539/66.

PT Low adenosine (A) content antisense oligonucleotides which do not trigger
 PT adenosine receptors during metabolism, useful e.g. for treating cancers
 PT and respiratory obstructions.

XX Disclosure; Page 1349-1353; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antistatic, hypotensive and cyostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced, specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
 CC surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention

SQ Sequence 17904 BP; 6014 A; 2945 C; 3330 G; 5615 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 3; Length 17904;

Best Local Similarity 81.5%; Pred. No. 93; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACACCTTAATTGACCGAGTACATTG 28

4640 ACACCTTAATTGACCGAGTACATTG 4614

RESULT 14
 AB297039/c

ID AB297039 standard; DNA; 17904 BP.

XX AB297039;

DT 17-OCT-2003 (first entry)

XX Human nucleic acid sequence.

XX Human; antisense; lung dysfunction; nasal airway dysfunction;
 KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
 KW antistatic; hypotensive; immunosuppressive; cyostatic; gene therapy;
 KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
 KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
 KW lung inflammation; respiratory disease; ds.

XX Homo sapiens.

XX WO200285308-A2.

XX 31-OCT-2002.

XX 23-APR-2002; 2002WO-US013135.

XX 24-APR-2001; 2001US-0286137P.

XX (EPIC-) EPIGENESIS PHARM INC.
 PA
 XX

PI Nyce JM, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D,
 PI Miller S, Tang L, Shahabuddin S;
 XX

DR WPI; 2003-229219/22.

XX Pharmaceutical composition for treating ailments associated with impaired
 PT respiration, has oligo(s) antisense to specific gene(s) or its
 PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
 PT ubiquinone.

PS Disclosure; SEQ ID NO 12281; 872pp; English.

XX The invention relates to a novel pharmaceutical composition, which has a
 CC first active agent comprising an oligonucleotide antisense to the
 CC initiation codon, coding region, 5' or 3' end genomic flanking regions,
 CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
 CC junctions of genes encoding a polypeptide associated with lung and/or
 CC nasal airway dysfunction and a second active agent comprising an
 CC antiinflammatory steroid and ubiquinone. A composition of the invention
 CC has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
 CC immunosuppressive, and cytostatic activity. The composition may have a
 CC use in antisense gene therapy. The composition is useful for treating or
 CC preventing a respiratory, lung or malignant disease or condition, also
 CC for enhancing the prophylactic or therapeutic respiratory effect of an
 CC antiinflammatory steroid in a subject, for reducing or depleting levels
 CC of, or reducing sensitivity to adenovirus, reducing levels of adenovirus
 CC receptor, producing bronchodilation, increasing levels of ubiquinone or
 CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
 CC lung inflammation, lung allergies, or a respiratory disease or condition.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 17904 BP; 6014 A; 2945 C; 3330 G; 5615 T; 0 U; 0 Other;

XX Query Match 63.3%; Score 19; DB 7; Length 17904;

XX Best Local Similarity 81.5%; Pred. No. 93;
 XX Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACACCTTAATTGACCGATACATTCG 28

DB 4640 ACACCTTAATTGACCGATACATTCG 4614

RESULT 15

AB067005/C
 ID AB067005 standard; DNA; 33053 BP.

AC AB067005;

DT 28-AUG-2002 (first entry)

XX Human angiogenesis associated polynucleotide SEQ ID NO 35.

XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
 KM inflammation; rheumatoid arthritis; diabetic retinopathy; antileucers;
 KM macular degeneration; inflammatory bowel disease; Crohn's disease;
 KM antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KM antarteriosclerotic; ds.

XX Homo sapiens.

XX WO200246454-A2.

XX 13-JUN-2002.

XX 06-DEC-2001; 2001WO-EP014320.

XX 06-DEC-2000; 2000DE-01061338.

PA (EPIC-) EPIGENOMICS AG.

XX Schächl O;

DR WPI; 2002-500450/53.

XX New nucleic acid fragments from chemically treated angiogenesis-
 PT associated genes, useful for determining methylation status, e.g. in
 PT diagnosis or treatment of cancer.

PS Claim 1; SEQ ID NO 35; 41pp + Sequence Listing; German.

XX The invention relates to a nucleic acid (I) comprising a segment of 18
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
 CC having sequences (AB066971-AB067178) or their complements. (I), also
 CC related oligomers, are used to evaluate the methylation status and/or
 CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
 CC diagnosis and treatment of eye diseases, proliferative retinopathy,
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
 CC Crohn's disease. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 33053 BP; 9206 A; 283 C; 6714 G; 16850 T; 0 U; 0 Other;

XX Query Match 63.3%; Score 19; DB 6; Length 33053;

XX Best Local Similarity 81.5%; Pred. No. 16+02;
 XX Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 CACTTAATTGACCGATACATTCGC 29

DB 7969 CACTTAATTGACCGATACATTCGC 7943

Search completed: July 20, 2004, 00:29:10
 Job time : 167.754 secs

This Page blank (uspioj)